SMALL RNA-MEDIATED REGULATION OF THE OXIDATIVE STRESS RESPONSE IN HALOARCHAEA

by
Diego Rivera Gelsinger

A dissertation submitted to Johns Hopkins University in conformity with the requirements for
the degree of Doctor of Philosophy

Baltimore, Maryland
March 2020

Abstract

Small RNAs (sRNAs), a type of non-coding RNA, are ubiquitously found in the three domains of life where they play large-scale roles in gene regulation. While a substantial body of experimental work has been done to uncover function of sRNAs in Bacteria and Eukarya, the functional roles of sRNAs in Archaea are still poorly understood. Archaea are often found in the most extreme environments on Earth, suggesting that they have evolved unique adaptions to respond to these stressful conditions. Archaea remain understudied in their regulatory processes of the central dogma, which is a hybrid system between Bacteria and Eukarya together with features unique to the Archaea. The goal of this thesis was to characterize sRNAs and elucidate their regulatory mechanisms in response to oxidative stress in the model extremophilic archaeon, *Haloferax volcanii*. Oxidative stress was investigated because it is a universal environmental stressor that produces robust phenotypes.

Using high-throughput RNA sequencing, I first characterized the transcriptional landscape of *H. volcanii* during oxidative stress. The oxidative stress response included upregulation of redox homeostasis, protein turnover, and DNA repair genes, and down regulation of metabolic genes. I also discovered thousands of novel sRNAs, comprising 25 to 30% of the total *H. volcanii* transcriptome, and found that hundreds of these sRNAs were differentially expressed during oxidative stress. Next, using ribosome profiling, I revealed the translational landscape, for the first time in *H. volcanii,* allowing for the validation of previously identified sRNAs. In addition, I discovered many novel translation start sites that may produce functionally important proteins during oxidative

stress. Third, I elucidated the mechanism of action of the most up-regulated sRNA during oxidative stress in *H. volcanii* and its functional role in survival. Finally, I developed and applied a new sRNA annotation pipeline to an extremophilic microbial community inhabiting halite (salt) rocks in the Atacama Desert, Chile. I discovered sRNAs expressed from diverse microorganisms and provided evidence for their functional relevance in the natural environment. The implications of this thesis work are that large-scale gene regulation is essential for Archaea to thrive in extreme environments and that sRNAs play key roles in those regulatory processes.

Primary Reader and Advisor: Dr. Jocelyne DiRuggiero

Secondary Reader: Dr. Sarah Woodson

Dedication

I dedicate this to the beautiful people who raised me as the third person of this family. To my mother, Carolina Rivera Escamilla, and my father, Charles Francis Gelsinger; you have been the foundation to all of this and will undoubtedly continue to be. I will always appreciate your love and look forward to our intricate conversations for years to come.

I dedicate this to my family. To my second mother, Blanca Rivera, who taught me compassion for not only my cousins – in all reality my brothers – but for this complex world. To my uncle, tio Rafael Escamilla, who has always fostered my creativity. To my cousins, Dorian, Kevin, and Emerson, who are my peers and help me grow every day. To my uncles and aunts – las tias y los tios – who helped me with whatever support they could muster in their busy lives. I love all of you.

Lastly, I want to dedicate this to all of my friends that have encouraged me and have helped me day to day in this process. The first, of course, has to be the stoop homies. Started in high school, and now we are here! Yervand Avanian, Elijah Navarro, and Pablo Goldstein, thank you for being my friends all this time. Next, my friends at SF State: Jon Kurland, Ricardo Portley, Austin Saladin, and Andrew Ward. You are all beloved friends who have helped in this journey in so many ways. To my best friends in graduate school; Brendan Miller and Michael Tisza. I wouldn't have survived graduate school without you both. To Emine Ertekin and Katie Farney who were my best friends

in lab and helped me not only excel in my work during the last year of my PhD, but to also have the most fun possible while doing it.

Acknowledgments

I first want to thank my thesis mentor, Dr. Jocelyne DiRuggiero. Without her mentorship, I would not be where I am today. She taught me how to be a great scientist but to also have compassion for others despite the hardships we face in academia. She provided me with guidance but also with the flexibility and independence I needed to thrive in academic research. She has opened the doors for me to network all around the world. I will try my hardest to repay this excellent mentorship forward, when I am a PI, to my own graduate students.

I also want to thank my undergraduate thesis mentor, Dr. José R. de la Torre. I asked to do research in his lab without knowing much about the world of academia at all and he took a chance on me. He introduced me to the world of biology, instilled in me the rigor necessary for excellent scientific research, but also showed me how to have fun and excitement with my work. He will always have a lasting impact on my scientific framework. He continued to be an important part of my graduate career through mentorship at conferences we both attended, and I anticipate our paths will continue to cross for years to come.

I thank the multitude of collaborators and mentors who have helped me tremendously throughout my years of research such as, but not limited to: Dr. Zhao Zhang, Dr. Evan Hass, Dr. Vidya Balagopal, Dr. Fuad Mohammad, Dr. Colin Wu, Dr. Kazuki Saito, Dr. Boris Zinshteyn, Dr. Niladri Sinha, Dr. Rachel Green, and Dr. Allen Buskirk. The majority of my thesis relied on next generation sequencing data, so I thank Dr. Jacques Ravel, Mike Humphrys, and David Mohr for sequencing efforts and technical advice. Lastly, I thank my thesis committee, Dr. Gisela Storz, Dr. John Kim and

Dr. Sarah Woodson for keeping me on track, for helpful discussions, and for sitting down and meeting with me whenever I was truly stuck on a difficult problem.

Contents

List of Tables

List of Figures

MAIN FIGURES

SUPPLEMENTAL FIGURES

Chapter 1

This work has been published in Genes - Gelsinger DR & DiRuggiero J, Genes March 2018, 5;9(3). pii: E141. doi: 10.3390/genes9030141

INTRODUCTION

Small RNAs (sRNAs) are important regulators for multiple cellular functions and they are ubiquitous in all domains of life. sRNAs, a class of non-coding RNAs (ncRNAs), are RNAs that do not function as messenger RNAs (mRNAs), ribosomal RNAs, or transfer RNAs in the cell. sRNAs in Bacteria and Eukarya play essential roles in transcriptional regulation, chromosome replication, RNA processing and modification, mRNA stability and translation, and even protein degradation and translocation (Cech & Steitz, 2014; Storz, Vogel, & Wassarman, 2011; Wagner & Romby, 2015). Recently, it was discovered that archaeal genomes encode for large numbers of sRNAs and that many of them are responsive to environmental stresses (Babski et al., 2016; Buddeweg et al., 2017; Fischer et al., 2011, 2010; Gelsinger & DiRuggiero, 2018; Heyer et al., 2012; Marchfelder et al., 2012; Prasse et al., 2011; Soppa et al., 2009; Straub et al., 2009).

While much remains to be elucidated about sRNAs in Archaea, decades of research in Eukarya and Bacteria have built a body of knowledge on their functional roles and their mechanisms of action. In Eukarya, several types of sRNAs have been identified, including microRNAs (miRNAs), PIWI-associated RNAs (piRNAs), and endogenous small interfering RNAs (siRNAs) (Cech & Steitz, 2014). The most studied of these, miRNAs, regulate protein expression in key cellular processes. miRNAs are typically 20-30

nucleotides (nt) long, target the 3' end of their cognate mRNAs, and form complexes with Argonaute (Ago) proteins (Cech & Steitz, 2014). While Ago homologs have also been found in archaeal genomes, there is no evidence for eukaryotic-like RNA interference in these organisms (Li et al., 2010). Rather, a defensive role against foreign genetic material was recently proposed whereby archaeal Ago proteins direct guide-dependent cleavage of foreign DNA (Zander et al., 2017).

In bacteria, sRNAs are typically 50 to 300 nt in length and act by targeting mRNA stability, translation, or by binding to proteins (Gottesman & Storz, 2011). Base-pairing with their mRNAs targets are of two types. *Cis*-encoded antisense RNAs (asRNAs) are encoded on the DNA strand opposite their target gene and thus can act via extensive base pairing (**Fig. 1-1A**). asRNAs have been found to repress transposons and toxic protein synthesis and to modulate the expression of transcriptional regulators (Gottesman & Storz, 2011; Storz et al., 2011). In contrast, *trans*-encoded sRNAs are encoded at genomic locations distinct from their target mRNAs, such as intergenic regions, and act via limited base pairing (Thomason & Storz, 2010) (**Fig. 1-1B**). These sRNAs bind at the 5’ end or 3' end of their target, either blocking ribosome binding and/or triggering degradation of target mRNAs via the endoribonuclease RNaseE (Gottesman & Storz, 2011). sRNAs can also activate translation by preventing the formation of inhibitory secondary structures and therefore increasing ribosome binding (Storz et al., 2011). *Trans*-encoded sRNAs can target multiple genes, including key transcription factors and regulators and, as a consequence, a single sRNA can modulate the expression of very large regulons (Gottesman & Storz, 2011; Morris & Mattick, 2014). A typical example of that is the sRNA OxyS involved in the oxidative stress response in *E. coli* (Altuvia et al., 1997). In most

gram-negative bacteria, the RNA binding protein Hfq is required for function and/or stability of the sRNAs (Storz et al., 2011).

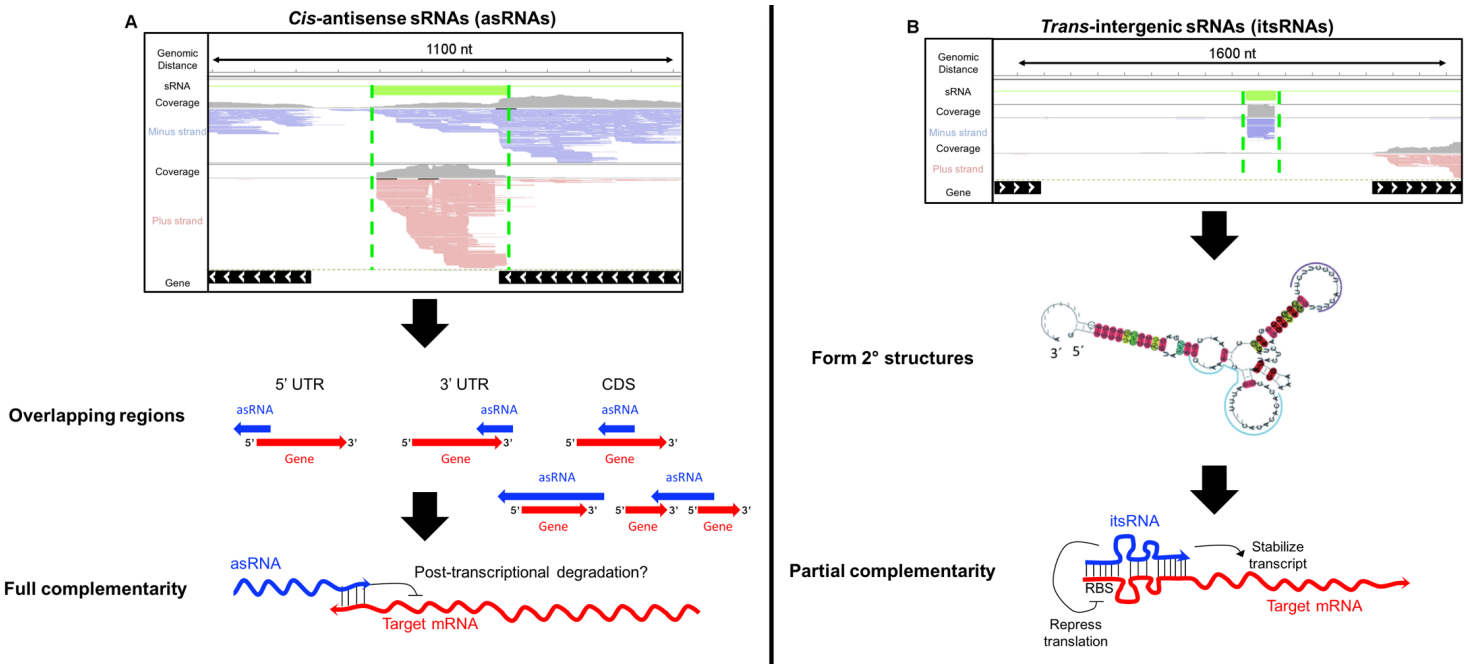

Figure 1-1: Classes of sRNAs discovered in *Archaea*. Genome viewer of (A) antisense sRNAs (*cis*-acting) and (B) intergenic sRNAs (*trans*-acting). Paired-end reads (100 bases) were mapped to the *H. volcanii* NCBI reference genome. Reference genes are marked as black lines with white arrows indicating their location on the plus strand (>) or minus strand (<). Reads marked in red are transcribed from the minus strand while blue reads are transcribed from the plus strand. Green lines indicated sRNAs. Coverage plots are in gray. Diagrams of classes of asRNAs are shown based on binding attributes: overlapping the 3' UTR, the 5' UTR, within the coding sequence (CDS), extending past the CDS, and overlapping multiple mRNAs. An example of an intergenic sRNA secondary structure is shown in (B) (Prasse et al., 2017) Reported regulatory roles of archaeal sRNAs are shown at the bottom.

Hfq is structurally related to the Sm/Lsm family of proteins and acts by stabilizing RNAs or by promoting rapid mRNA-sRNA base-pairing and recruiting of RNAseE for degradation (Storz et al., 2011). However, other bacteria do not require Hfq for sRNA-mediated regulation even when the protein is encoded in their genome. Recently, novel RNA-binding proteins such as CsrA and ProQ have also been proposed to function as sRNAs chaperones in bacteria (Smirnov et al., 2016).

In archaea, the functional and mechanistic characterization of sRNAs is still in its infancy. The size range of archaeal sRNAs is 50 to 500 nt in length. Both *cis* asRNAs and *trans*-encoded sRNAs (thereafter called itsRNAs for intergenic sRNAs) have been reported in a number of archaeal species (**Fig. 1-1**), as well as *cis* sense sRNAs that are transcribed within the open reading frame (ORF) of genes (Babski et al., 2016; Bernick et al., 2012; Gelsinger & DiRuggiero, 2018; Jäger et al., 2014; Jager et al., 2009; Li et al., 2015; Toffano-Nioche et al., 2013; Wurtzel et al., 2010). In addition to these ncRNAs, small nucleolar RNAs (snoRNAs), tRNA-derived fragments, and CRISPR RNAs (crRNAs) have been found in Archaea. This chapter will focus on the *cis*- and *trans*-encoded archaeal sRNAs as it becomes more evident that they play essential roles in gene regulation and because there have been exciting new developments in the last few years (since the last sRNA review) in unraveling the functional roles of these sRNAs. We will first document the sRNAs identified so far in the Archaea and discuss the state-of-the art methods for sRNA detection. We will then describe the molecular mechanisms that have been elucidated for a small number of archaeal sRNAs, give an overview of sRNA-interacting partners, and as such provide insights into the biological roles of these sRNAs. Lastly, I

will introduce the overall objective of my thesis work and my specific aims of my dissertation.

Identification of sRNAs: what has been found so far?

In contrast to the wealth of knowledge on bacterial and eukaryal RNA regulators, our knowledge of sRNAs in Archaea is limited to a handful of studies for hyperthermophiles, methanogens, and the haloarchaea (*Haloferax volcanii* and *Haloferax mediterranii*) (Laass et al., 2019; Marchfelder et al., 2012; Payá et al., 2018; Prasse et al., 2013; Soppa et al., 2009; Wyss et al., 2018). Both classes of trans- and cis-encoded regulatory sRNAs have been found in the Archaea. Initial identification of sRNAs relied on bioinformatic (RNomics) prediction using archaeal whole genome sequences. Later, microarray and 454-pyrosequencing technologies provided a mean to validate these predicted sRNAs and further identified novel sRNAs by the hundreds. However, it is the unprecedented discovery of more than 2,900 sRNAs in *H. volcanii* by multiple recent high-throughput sequencing (HTS) studies - which is quite remarkable considering that the genome of this organism encode for just over 4,000 proteins (Babski et al., 2016; Gelsinger & DiRuggiero, 2018; Laass et al., 2019; Wyss et al., 2018) - that has permanently altered our view of the archaeal transcriptome. From these studies it is now clear that a large proportion of RNAs are non-coding, nearly rivaling the number of RNAs encoding for proteins (Babski et al., 2016; Gelsinger & DiRuggiero, 2018; Laass et al., 2019; Wyss et al., 2018). Additionally, both non-coding RNAs up to 1,000 nt in length and non-coding RNAs that associate with ribosomes (rancRNAs) were also reported (Gelsinger & DiRuggiero, 2018; Wyss et al., 2018), thus increasing the size range and regulatory potential of identified archaeal non-coding RNAs. However, because of the deep level of sequencing in HTS studies it is

important to use thresholding criteria to distinguish small RNAs from transcriptional noise or non-functional transcripts. Such criteria may include the presence of promoter elements, conservation across taxa, minimum expression levels, and differential regulation under specific conditions.

In the model haloarchaeon, *H. volcanii*, as many as 1,500, asRNAs and 400 intergenic sRNAs have been identified, indicating that most sRNAs in this organism are antisense to coding regions. Furthermore, as much as 30% of the sRNAs discovered in *H. volcanii* contained stringent basal transcriptional promoters, such as a TATA-box, and exhibited expression levels comparable to mRNAs, underling their relevance in the global regulation of gene networks (Babski et al., 2016; Gelsinger & DiRuggiero, 2018; Laass et al., 2019).

While sRNAs are particularly numerous in haloarchaea genomes (Babski et al., 2016), they have also been found in a number of other archaea, including *Sulfolobus* (Wurtzel et al., 2010), *Methanosarcina* (Jager et al., 2009), *Pyrobaculum* (Bernick et al., 2012), *Pyrococcus* (Toffano-Nioche et al., 2013), *Thermococcus* (Jäger et al., 2014), and *Methanolobus* (J. Li et al., 2015) (**Table 1**). In *Sulfolobus solfataricus*, 125 trans-encoded sRNAs and 185 cis-encoded asRNAs were identified using HTS, suggesting that 6.1% of all genes in *S. solfataricus* are associated with sRNAs (Wurtzel et al., 2010). A comparative genome analysis of *Methanosarcina mazei*, *M. bakeri*, and *M. acetivorans* revealed that 30% of the antisense and 21% of the intergenic sRNAs identified were conserved across the 3 species (Jager et al., 2009). The number of antisense sRNAs reported in the archaea numbers in the hundreds and further work is needed to validate and characterize their functional roles (Thomason & Storz, 2010). It should be noted that of some the differences in numbers of sRNAs in **Table 1** may be attributed to differences in sequencing

technologies and sequencing depths used in these studies. In particular, studies that used microarray versus 454-sequencing and HTS with Illumina platforms, the latter two allowing for *de novo* discovery of the entire transcriptome. The sRNA numbers reported in Laass, *et al*. (2019), Wyss, *et al.* (2018), Gelsinger, *et al.* (2018), Babski, *et al.* (2016), Li, *et al.* (2015), Jäger, *et al.* (2014), and Toffano-Nioche, *et al.* (2013) all used HTS Illumina technologies and are thus most comparable with each other. All other studies used microarray or 454-sequencing technologies.

	Number of Genes	Total Number of sRNAs	Number of itsRNAs	Number of asRNAs	Number of iRNAs	Reference
Euryarchaeota						
Haloferax volcanii	4023	1635	200	1222	213	Laass, *et al.* (2019)
Haloferax volcanii	4023	1169	245	68	856	Wyss, *et al.* (2018)
Haloferax mediterranii	3884	57	51	6	N/A	Payá, *et al.* (2018)
Haloferax volcanii	4023	1557	77	1480	N/A	Gelsinger, *et al.* (2018)
Haloferax volcanii	4023	2792	395	1244	1153	Babski, *et al.* (2016)
Haloferax volcanii	4023	190	145	45	N/A	Heyer, *et al.* (2012)
Methanolobus psychrophilus	2974	2745	195	1110	1440	Li, *et al.* (2015)
Methanosarcina mazei	3551	242	199	43	N/A	Jäger, *et al.* (2009)
Thermococcus kodakarensis	2328	1731	69	1018	644	Jäger, *et al.* (2014)
Pyrococcus abyssi	1969	322	107	215	N/A	Toffano-Nioche, *et al.* (2013)
Archaeoglobus fulgidus	248	45	9	33	3	Tang, *et al.* (2002)
Crenarchaeaota						
Sulfolobus solfataricus	3254	310	125	185	N/A	Wurtzel, *et al.* (2010)
Pyrobaculum aerophilum, arsenaticum, calidifontis, & islandicum	2706, 2407, 2200, 2075	Number Not Reported	Number Not Reported	3	N/A	Lowe, *et al.* (2012)
Nanoarchaeaota						
Nanoarchaeum equitans	553	Number Not Reported	Number Not Reported	Number Not Reported	N/A	Randau. (2012)
asRNAs: antisense sRNAs; itsRNAs: intergenic sRNAs; iRNAs: internally transcribed sRNAs; N/A: these types of sRNAs were not reported.						

Table 1: Summary of sRNA discovered in the *Archaea*

Best methods for sRNA discovery

Methods currently used to discover sRNAs in Archaea are all RNA sequencing methods that take advantage of the sequencing depth and high throughput of Illumina technologies. These methods are (1) differential RNA-sequencing (dRNA-seq) and (2) size-selected, strand-specific sRNA-sequencing (sRNA-seq) (Babski et al., 2016; Crits-Christoph et al., 2016).

Differential RNA-seq was used to identify hundreds to thousands of sRNAs in *H. volcanii*, *M. psychrophilus*, *T. kodakerensis*, and *P. abyssi* (Babski et al., 2016; Jäger et al., 2014; J. Li et al., 2015; Toffano-Nioche et al., 2013). The dRNA-seq method is based on the selective enrichment of primary transcripts, and allows for transcription start site mapping (Sharma & Vogel, 2014). This provides a global approach to identify all primary RNAs and the exact position at which they are transcribed, under any condition (Sharma & Vogel, 2014). However, a significant drawback to this method is that it does not provide information on the length of the sRNAs because it is restricted to the 5′-ends of transcripts; it is also biased against processed sRNAs (Babski et al., 2016). Another method for sRNA identification is presented in Gelsinger *et al.* (2018) and uses a modified sRNA-seq protocol that enabled strand-specific deep sequencing and identification of thousands of sRNAs in *H. volcanii*. In this method, RNA is size-selected and strand-specificity is preserved (Gelsinger & DiRuggiero, 2018). By significantly enriching for sRNAs, this method provides better detection of full length of sRNAs and its strand specificity allowed for the clear identification of antisense and intergenic sRNAs, and potential targets of these sRNAs. However, the detection of internal sense sRNAs appeared to be difficult because of their masking by mRNA reads.

Besides library preparation and sequencing, another major difficulty in sRNA identification is in the bioinformatic analysis of the RNA-seq data. While no single pipeline has been published to specifically identify sRNA in Archaea, the computational strategy used in Gelsinger et al. (2018) presented a significant step forward in designing an analytical pipeline specific for archaeal sRNA discovery. This pipeline was then adapted into a program that can use both isolate transcriptomic and community metatranscriptomic RNA-seq data to identify sRNAs in both Archaea and Bacteria, named SnapT (Gelsinger et al., 2020).

Molecular regulatory mechanisms and targets of sRNA in Archaea

A. Antisense sRNAs (asRNAs)

Despite the discovery of thousands of sRNAs in archaeal transcriptomes, functional and mechanistic characterizations of sRNAs in the *Archaea* is in its infancy. Initial insight into antisense sRNA mechanisms comes from recent work in *H. volcanii,* showing that an overwhelming majority of all sRNAs expressed in this organism are antisense (Babski et al., 2016; Gelsinger & DiRuggiero, 2018; Laass et al., 2019) (**Fig. 1-2**). Of these, only a minority (7%) overlapped the 5' UTR of mRNAs, which is in concurrence with findings that most mRNAs in *H. volcanii* are leaderless (lacking a 5' UTR) (Gelsinger & DiRuggiero, 2018), while most (67%) overlapped within the coding sequence (CDS) of mRNAs (**Fig. 1-2**). In bacterial itsRNAs and eukaryal sRNAs, the region of interaction (hybridization) between a sRNA and its target mRNA has been termed a "seed" region (Cech & Steitz, 2014). In *H. volcanii* no "seed" binding region for CDS-binding sRNAs had been found, indicating that they could potentially hybridize to the entire length of the

mRNA (full occupancy) (Gelsinger & DiRuggiero, 2018). However, while thermodynamically favorable, the hybridization across the full length of CDS-asRNAs has not been demonstrated *in vivo*. Additionally, some of these CDS-asRNAs have the potential to form secondary structure, suggesting that the CDS-asRNA might only binds part of the transcript (a partial "seed" region). Lastly, a smaller fraction (26%) of asRNAs overlapped the 3' UTR of mRNAs (Gelsinger & DiRuggiero, 2018) **(Fig. 1-2)**.

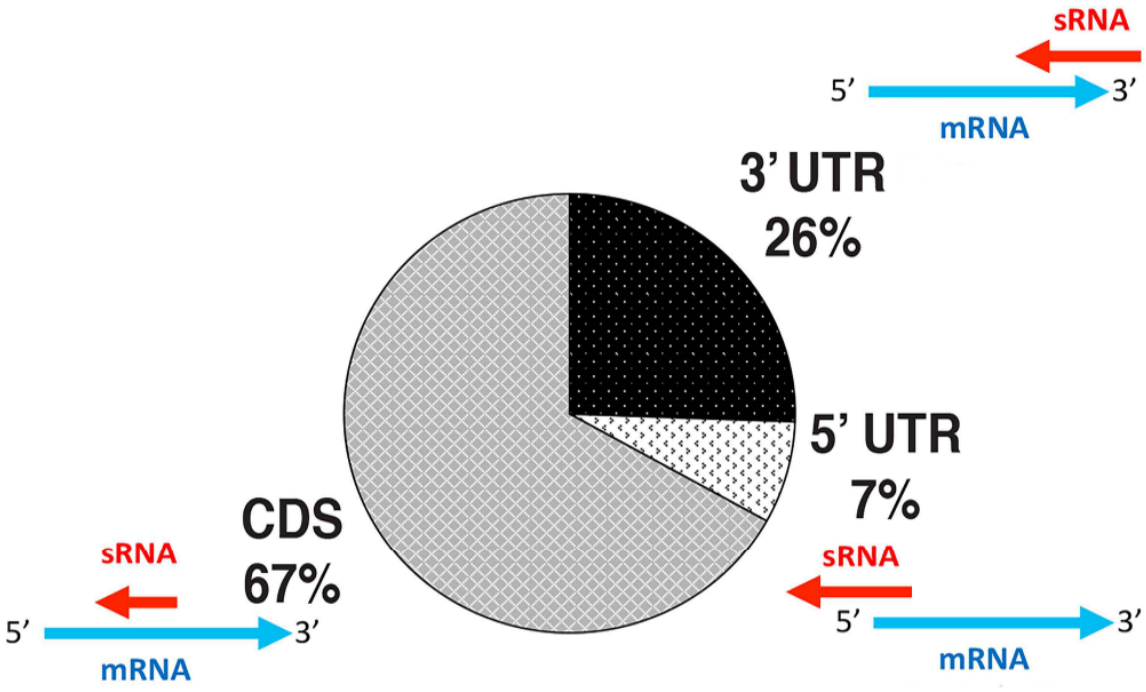

Figure 1-2: Distribution of binding regions for antisense sRNAs. UTR, untranslated region; CDS, coding sequence (Gelsinger & DiRuggiero, 2018).

There are many known advantages of sRNA regulators including reduced metabolic cost, additional levels of regulation, unique regulatory properties, and faster response to stresses. Indeed the regulatory effects of sRNAs are often observed within minutes in bacterial systems (Storz et al., 2011). Furthermore, sRNAs in bacteria can regulate very large gene networks as well as key transcription factors (Storz et al., 2011).

Examples of these include OxyS and SgrS in *E. coli,* involved in oxidative and glucose-phosphate stress, respectively (Altuvia et al., 1997). Antisense sRNAs, which are by far the largest group of sRNAs found in the *Archaea*, are encoded in the opposite strand of their putative target. In the hyperthermophile *Pyrobaculum*, three antisense sRNAs were found opposite a ferric uptake regulator, a triose-phosphate isomerase, and transcription factor B, supporting a potential role in the regulation of iron, transcription, and core metabolism (Bernick et al., 2012). Target enrichment of asRNAs differentially regulated by oxidative stress in *H. volcanii* included mRNAs involved in transposon mobility, chemotaxis signaling, peptidase activity, and transcription factors (Gelsinger & DiRuggiero, 2018). The functional enrichment of transposon targeted by asRNAs suggests that during oxidative stress transposon activity is tightly regulated in *H. volcanii*, potentially explaining its increased resistance to oxidative stress conditions (Gelsinger & DiRuggiero, 2018). Indeed, transposons are genetic elements that hop around in the genome causing double strand breaks. This added stress would likely be detrimental to a cell under oxidative stress, hence a need to be silenced (Whitehead et al., 2006). sRNAs antisense to transposons were also reported for *Thermococcus kodakarensis* (Jäger et al., 2014), *S. solfataricus* (Wurtzel et al., 2010), and *M. mazei* (Jager et al., 2009) suggesting that, similarly to bacteria, regulation of transposition is mediated by asRNAs in archaea (Balasubramanian, Ragunathan, Fei, & Vanderpool, 2016). Initial mechanistic insight of asRNAs comes from a recent study of sRNAs in *H. volcanii* by Gelsinger, *et al* 2018. which found that a large number of asRNAs were either up-regulated or down-regulated during oxidative stress, revealing two types of antisense sRNA populations. An anti-correlation was observed for a group of up-regulated antisense sRNAs and their down-

regulated *cis*-encoded putative targets, indicating a potential mechanism of negative regulation (Gelsinger & DiRuggiero, 2018). In contrast, many *cis*-encoded putative mRNA targets and their cognate asRNAs exhibited a positive correlation in their expression patterns to oxidative stress, suggesting a positive regulatory effect between asRNA-mRNA *cis*-pairs (Gelsinger & DiRuggiero, 2018). Although negative regulatory effect of asRNAs on their target mRNAs was also suggested in another study, also in *H. volcanii* (Babski et al., 2016), experimental evidence are still lacking because of the inherent difficulty at manipulating such overlapping sRNA-mRNAs pairs.

B. Intergenic sRNAs (itsRNAs)

While the regulatory effects of asRNAs can be readily inferred because of the overlap with their mRNAs targets, it is rather different with intergenic sRNAs where finding targets is a particularly difficult task. As a consequence, mechanistic insights into the regulation of intergenic sRNAs have only been provided for very few specific sRNAs. In *H. volcanii*, many intergenic sRNAs are differentially expressed in response to varying environmental conditions, including different growth states (exponential vs stationary), alternative sugar sources, elevated temperature, osmotic stress, nutrient limitation, and oxidative stress (J Babski et al., 2011; Gelsinger & DiRuggiero, 2018; Heyer et al., 2012; Jaschinski et al., 2014; Laass et al., 2019; Straub et al., 2009). While phenotypic characterization of sRNA deletion mutants, including 10 gain-of-function phenotypes out of 27 mutants tested, confirmed their roles in metabolic regulation, stress adaptation and complex behavior (Jaschinski et al., 2014; Straub et al., 2009), their targets are still unknown with a few exceptions.

Some of these exceptions came from the study of *M. mazei* cultures grown under nitrogen starvation conditions where RNA-seq experiments revealed the differential expression of a number of sRNAs in response to nitrogen availability (Jager et al., 2012, 2009). This then resulted in the identification of the first *in vivo* targets for archaeal intergenic sRNAs (Jager et al., 2012, 2009). A potential target for one of these sRNAs, $sRNA_{162}$, was a bicistronic mRNA encoding for a transcription factor involved in regulating the switch between carbon sources and for a protein of unknown function (Jager et al., 2012). Another sRNA, $sRNA_{154}$, was also exclusively expressed during nitrogen starvation conditions and the multiple targets for $sRNA_{154}$ included mRNAs for the α subunit of nitrogenase (*nif*H), the transcriptional activator of the *nif* operon (*nrp*A), and glutamine synthase1/2 (*gln*A_1/*gln*A_2). Prasse et al. (Daniela Prasse et al., 2017) determined that $sRNA_{154}$ stabilized some mRNAs while inhibiting translation initiation for other mRNAs, thus playing a dual regulatory role (**Fig. 3**). $sRNA_{154}$ was found to stabilize nifH, nrpA-, and $glnA_1$-mRNAs but to block the translation of $glnA_2$-mRNA. $sRNA_{154}$ is highly conserved in the *Methanosarcina* and it was predicted to form a stable secondary structure with two loops required for stabilization of mRNA targets. The authors of the study proposed that the mechanism of the two loops was to mask endonucleolytic cleavage sites of RNases by hybridizing to the mRNA targets and preserving the mRNA for translation (Daniela Prasse et al., 2017). In contrast, they also showed that loop 2 of $sRNA_{154}$ contains anti-ribosome binding site (RBS) sequences that masked the RBS of the *glnA*$_2$-mRNA target, repressing translation initiation (Daniela Prasse et al., 2017). The proposed functional role of $sRNA_{154}$ was to regulate N_2-fixation under nitrogen limiting conditions by stabilizing transcripts involved in nitrogenase production (both regulators of and the

nitrogenase itself), leading to a feed forward regulatory system (Daniela Prasse et al., 2017).

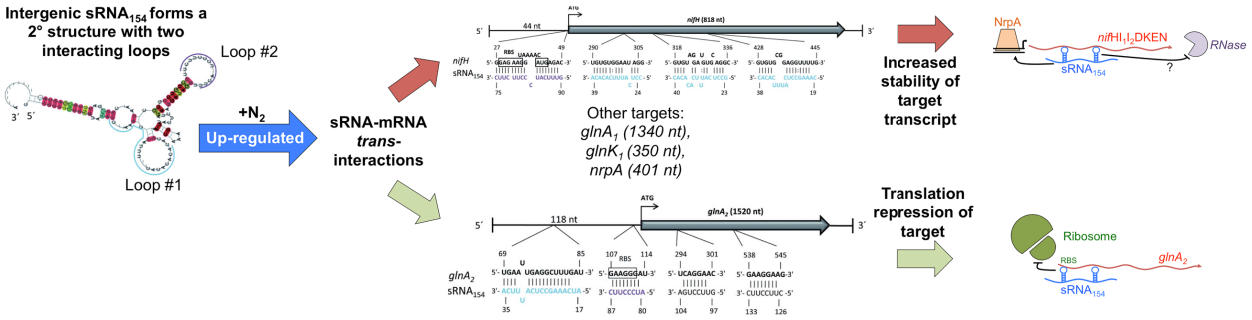

Figure 1-3: Two proposed mechanisms of action and targets for the itsRNA, $sRNA_{154}$, in *M. mazei* under conditions in which N_2 is the only source of nitrogen. $sRNA_{154}$ is predicted to form a stable secondary structure with two stem loops that interact with mRNA targets. Mechanism 1 (top): target stabilization by sRNA-mediated masking of sites for degradation by an unknown RNase for the *nif* operon and the transcription factor *nrp*A, which regulate the *nif* operon. Mechanism 2 (bottom): $sRNA_{154}$ binds to the $glnA_2$ ribosome binding site (RBS) via loop #2, preventing translating and thus decreasing the amount of protein produced but not that of the transcript.

Most recently, another *M. mazei* itsRNA, $sRNA_{41}$, was found to be down-regulated during nitrogen limiting conditions. Targets of $sRNA_{41}$ were involved in acetyl-CoA- decarbonylase/synthase complexes (ACDS) and were repressed at the translational level (Buddeweg et al., 2017). Thus $sRNA_{41}$ was predicted to play a role in repressing ACDS protein levels, however, under nitrogen limiting conditions, $sRNA_{41}$ was down-regulated, thus allowing ACDS levels to increase, which in turn provided sufficient amino acids for nitrogenase synthesis and energy for N_2-fixation (Buddeweg et al., 2017). Thus the proposed mechanism of regulation for $sRNA_{41}$ was to repress its targets at the translation level by masking ribosome binding sites within polycistronic mRNAs

(Buddeweg et al., 2017). Other molecular mechanisms have been identified in the archaea such as the binding of itsRNAs to the 3'UTR of mRNA targets in *S. solfataricus* (Märtens et al., 2013) and, more recently, in *H. volcanii* (Kliemt *et al*., unpublished data). This is of particular interest in the haloarchaea because 72% of their transcripts are leaderless (Babski et al., 2016).

While these studies provide great examples of gene network regulated by sRNAs in the archaea, additional work is needed to identify many more molecular targets of archaeal itsRNAs and the diverse mechanisms of their sRNA-mRNA interactions. Taken together these studies are building a narrative of sRNA (both antisense and intergenic) mechanisms in the archaea, combining global approaches with individual targeted sRNA studies, demonstrating that sRNAs are also essential partners in gene regulation in the third domain of life **(Fig. 1-4)**. Furthermore, each of these studies have shown that sRNAs are abundant in Archaea and have the potential to play large-scale roles in gene regulation in response to a variety of environmental stresses.

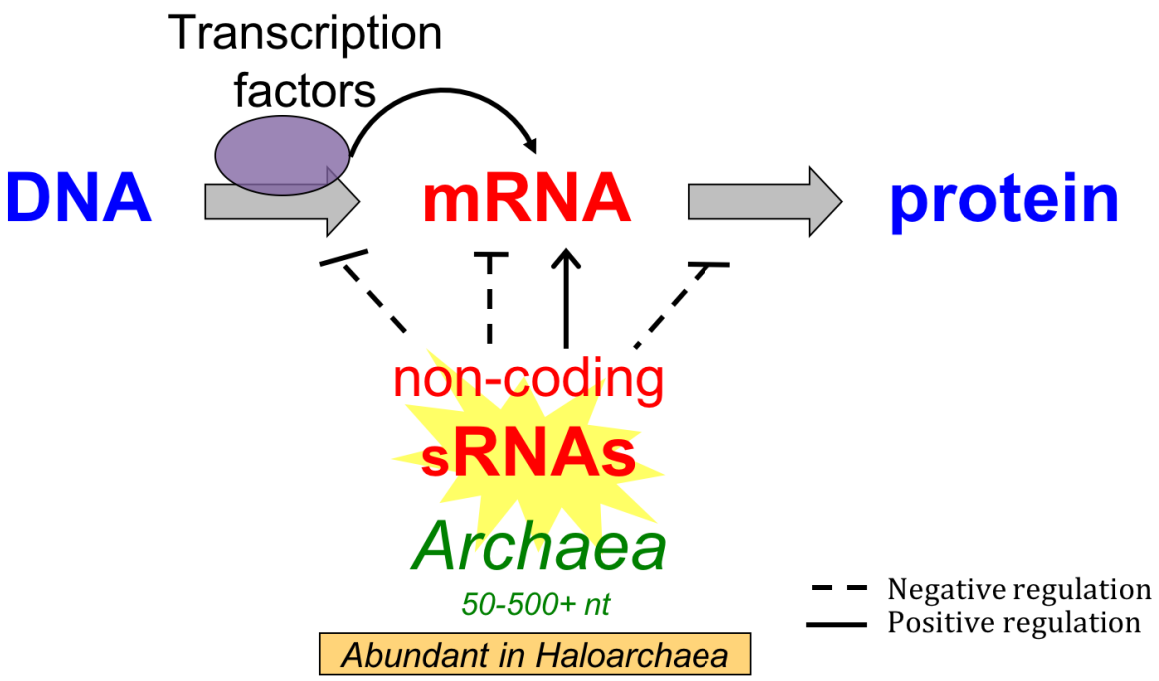

Figure 1-4: Correlation- and experimental-based regulatory mechanisms for sRNAs reported in the *Archaea*. The grey arrows indication the processes of transcription and translation. Regulatory arrows indicate regulation at transcription, post-transcription, and translation steps.

Overall objective and specific aims

In this dissertation, I used modern techniques to shed new light on an old question: how do extremophilic archaea adapt to some of the most inhospitable environments on Earth? The Archaea are of special interest because they stand at the crossroad between the other two domains of life by using a mosaic of molecular features from both Bacteria and Eukarya, along with unique features. For example, Archaea have a eukaryotic-like transcription apparatus (i.e., RNA-pol and initiation) but bacterial-like

regulation and coupled transcription-translation. Here, I postulate that a key factor to adaptation to extreme environments is highly coordinated gene regulation.

As described above, sRNAs are ubiquitously found in the three domains of life, playing large-scale roles in gene regulation, transposable element silencing and defense against foreign elements. While a substantial body of experimental work has been done to uncover function of sRNAs in Bacteria and Eukarya, the functional roles of sRNAs in Archaea are still poorly understood. The objective of this thesis was to elucidate the regulatory mechanisms of sRNAs in the oxidative stress response of a model archaeon, the halophile *Haloferax volcanii*. ***Thus, in my dissertation, I hypothesized that sRNA-mediated post-transcriptional regulation is a key stress response pathway in haloarchaea to environmental challenges and plays a role in their increased resistance to oxidative stress.*** A number of halophilic archaea have been developed into model systems for the study of biological processes. Objectively, the best haloarchaeal system is *Haloferax volcanii* with previously developed molecular tools that include highly efficient genetic systems, reporter genes, and biochemical methods (Pohlschroder & Schulze, 2019).

My fours dissertation aims are:

1. ***Probe the transcriptional landscape and regulatory roles of small noncoding RNAs in the oxidative stress response of the haloarchaeon H. volcanii***
    - To carry out this aim, I elucidated the oxidative stress response of *H. volcanii* at the transcriptional level by HTS mRNA-seq and differential expression analysis relative to no challenge conditions. Using strand-specific sRNA-seq, and a bioinformatic pipeline specifically optimized for

Archaea, I then discovered thousands of sRNAs in the transcriptome of *H. volcanii*, under no challenge and oxidative stress conditions. I also identified hundreds of differentially expressed sRNAs in response to hydrogen peroxide-induced oxidative stress in this organism.

2. ***Elucidate, using ribosome profiling, the translational landscape and alternative reading frame usage in Archaea***
    - Here, I probed the translation landscape of *H. volcanii* by developing ribosome profiling, an HTS technique that measures the number of ribosomes that are translating mRNA *in vivo* in the cell. I determined that the ribosome footprint of *H. volcanii* was comparable in size to that of Eukarya, supporting the close evolutionary relationship between the two domains. I used the drug, harringtonine, to arrest ribosomes at initiation sites of genes. This method allowed me to map known translation initiation sites of previously annotated genes across the genome, and also to reveal novel initiation sites in intergenic regions, on non-coding RNAs, and internally within known genes. I explored the translation of these alternative translation start sites (aTSS) during oxidative stress.
3. ***Determine the mechanism and function of the intergenic oxidative stress-specific sRNA, SHOxi, at phenotypic and molecular resolution.***
    - In this aim, using reverse molecular genetics on the most up-regulated sRNA, SHOxi, I assessed phenotypic and molecular changes to achieve functional understanding of the role of this sRNA in the oxidative stress response of *H. volcanii*. This approach included evaluating survival

defects and changes in transcription (RNA-seq and qPCR) and translation (ribosome profiling) changes in the absence of SHOxi.

4. ***Identify regulatory non-coding small RNAs in an extremophilic microbial community***

    - Lastly, I developed and applied a new sRNA annotation pipeline to an extremophilic microbial community abundant in haloarchaea in order to assess the functional relevance of sRNAs in a natural environment.

Chapter 2

Transcriptional Landscape and Regulatory Roles of Small Noncoding RNAs in the Oxidative Stress Response of the Haloarchaeon *Haloferax volcanii*

This work has been published in the Journal of Bacteriology – Gelsinger DR & DiRuggiero J, Journal of Bacteriology April 2018, 200 (9) e00779-17; DOI: 10.1128/JB.00779-17

ABSTRACT

Haloarchaea in their natural environment are exposed to hyper-salinity, intense solar radiation, and desiccation, all of which generate high levels of oxidative stress. Previous work has shown that *Haloarchaea* are an order of magnitude more resistant to oxidative stress than most mesophilic organisms. Despite this resistance, the pathways *Haloarchaea* use to respond to oxidative stress damage are similar to that of non-resistant organisms suggesting that regulatory processes might be key to their robustness. Recently, small non-coding RNAs (sRNAs) were discovered in *Archaea* under a variety of environmental conditions. We report here the transcriptional landscape and functional roles of sRNAs in the regulation of the oxidative stress response of the model haloarchaeon *Haloferax volcanii*. Thousands of sRNAs, both intergenic and antisense, were discovered using strand-specific sRNA-seq, comprising 25 to 30% of the total transcriptome during no-challenge and oxidative stress conditions, respectively. We identified hundreds of differentially expressed sRNAs in response to hydrogen peroxide induced oxidative stress in *H. volcanii*. Targets of antisense sRNAs decreased in expression when sRNAs were up-

regulated indicating that sRNAs are likely playing a negative regulatory role on mRNA targets at the transcript level. Target enrichment of these antisense sRNAs included mRNAs involved in transposons mobility, chemotaxis signaling, peptidase activity, and transcription factors.

IMPORTANCE While a substantial body of experimental work has been done to uncover functions of sRNAs in gene regulation in Bacteria and Eukarya, the functional roles of sRNAs in Archaea are still poorly understood. This study is the first to establish the regulatory effects of sRNAs on mRNAs during the oxidative stress response in the haloarchaeon *Haloferax volcanii*. Our work demonstrates that common principles for the response to a major cellular stress exist across the 3 domains of life while uncovering pathways that might be specific to the Archaea. This work also underscores the relevance of sRNAs in adaptation to extreme environmental conditions.

INTRODUCTION

Microbial communities that reside inside halite nodules from Salars in the Atacama Desert, Chile, are under extreme environmental pressures due to hyper-salinity, intense solar radiation, and frequent desiccation-hydration cycles, which all generate high levels of oxidative stress (Kaur et al., 2010; Robinson et al., 2015). Oxidative stress occurs when the level of reactive oxygen species (ROS) produced in cells overwhelms antioxidant defense mechanisms and damage accumulates (Imlay, 2008). Through metagenomic studies, we found the dominant populations in these salt rocks to be *Haloarchaea* such as *Haloferax* and *Halobacterium* (Crits-Christoph et al., 2016). These halophilic

microorganisms are members of the third domain of life, the *Archaea*. *Haloarchaea* have previously been shown to be highly resistant to ROS damage, withstanding many times what *E. coli* and other radiation-sensitive organisms can survive (C. K. Robinson et al., 2011; Sharma et al., 2012; Whitehead et al., 2006). The haloarcheon *H. salinarum* has been shown to use a wide-range of strategies to combat damage from oxidative stress including multiple copies of genomes (polyploidy) as substrate for DNA repair, functional redundancy of DNA repair and detoxification enzymes (e.g. catalase), increased cytosolic manganese complexes to scavenge ROS, and differential regulation of genes in response to stress (Sharma et al., 2017; C. K. Robinson et al., 2011; Sharma et al., 2012; Webb et al., 2013; Whitehead et al., 2006). However, pathways for DNA repair and protein turnover in *Haloarchaea* are nearly identical to non-resistant bacteria and eukarya suggesting that the regulation of these processes in response to oxidative stress might be key to their robustness. Previous work with *H. salinarum* oxidative stress gene regulatory networks revealed that a single transcription factor, RosR, regulates the appropriate dynamic response of nearly 300 genes to reactive oxygen species stress (Sharma et al., 2012). This work demonstrated that the oxidative stress response in *H. salinarum* impacted a wide array of cellular processes, engaging at least 50% of all the genes (Kaur et al., 2010). These results underline the importance of gene regulation in *Haloarchaea* for responding to and counteracting the damage caused by oxidative stress.

In addition to transcription factors, small regulatory RNAs (sRNAs) similarly act as global gene regulators (Morris KV, 2014). Small RNAs (sRNAs) are ubiquitously found in *Bacteria* and *Eukarya*, playing large-scale roles in gene regulation, transposable element silencing, defense against disease state, and foreign elements (Altuvia et al., 1997; Cech &

Steitz, 2014; Marchfelder et al., 2012; D. Prasse et al., 2013). Several types of sRNAs have been identified in the *Eukarya* (miRNAs, siRNAs, and piRNAs) and they are typically 20-25 nucleotides (nt) long. Their major mode of interaction is through base pairing to the 3'-unstranslated region (UTR) of their target mRNAs, inhibiting translation or triggering target degradation with associated protein components (Argonautes) (Morris KV, 2014). Bacterial sRNAs have been shown to modulate core metabolic functions and stress related responses, such as nutrient deprivation, by binding target mRNAs and causing their degradation or preventing translation (Altuvia et al., 2004; Altuvia et al., 1997). Most of the functionally characterized sRNAs in *Bacteria* bind the 5'-UTR of their target mRNA and are longer than their eukaryal counterparts, with sizes ranging from 50 to 500 nt. These sRNAs can target multiple genes, including key transcription factors and regulators (Altuvia et al., 2004; Altuvia et al., 1997; Gottesman & Storz, 2011). As a consequence, a single sRNA can modulate the expression of large regulons and thus have a significant effect on metabolic processes. For example, the bacterial sRNA OxyS, which is dramatically induced by oxidative stress, regulates the expression of about 40 genes and interacts directly with eight target mRNAs (Altuvia et al., 1997).

sRNAs have been discovered to be abundant in *Archaea*, more specifically in *Haloarchaea,* in response to a variety of environmental conditions but the functional roles of these RNAs still remain poorly understood nor has a link between sRNAs expression and oxidative stress response been established (Babski et al., 2016; Fischer et al., 2011; Fischer et al., 2010; Heyer et al., 2012; Marchfelder et al., 2012; Daniela et al., 2017; Schmitz-Streit et al., 2011; Soppa et al., 2009; Straub et al., 2009). Only a handful of studies on sRNAs in hyperthermophiles, methanogens, and the haloarchaeon *Haloferax volcanii*

have been reported so far (Babski et al., 2016; Fischer et al., 2011; Fischer et al., 2010; Heyer, Dorr, et al., 2012; Marchfelder et al., 2012; Prasse et al., 2017; Schmitz-Streit et al., 2011; Soppa et al., 2009; Straub et al., 2009). In *H. volcanii* a large number of intergenic- and antisense-encoded sRNAs, 145 and 45, respectively, were discovered using microarray in addition to a novel class of sRNAs recently described in eukaryotes, tRNA-derived fragments (tRFs), and a new study found thousands of sRNAs present in this organism (Babski et al., 2016; Gebetsberger et al., 2012; Heyer et al., 2012). In *Sulfolobus solfataricus*, 125 trans-encoded sRNAs and 185 cis-antisense sRNAs were identified using high-throughput sequencing (HTS), suggesting that 6.1% of all genes in *S. solfataricus* are associated with sRNAs (Wurtzel et al., 2010). A comparative genome analysis of *Methanosarcina mazei*, *M. bakeri*, and *M. acetivorans* revealed that 30% of the antisense and 21% of the intergenic sRNAs identified were conserved across the 3 species (Jager et al., 2009). Co-immuno-precipitation with the Lsm protein (archaeal Hfq homolog) was used to "capture" sRNAs (Fischer et al., 2010) but its functional role remains to be elucidated. While Ago homologs have also been found in archaeal genomes, there is no evidence for eukaryotic-like RNA interference in these organisms (Li et al., 2010). Rather, a defensive role against foreign genetic material was recently proposed whereby archaeal Ago proteins direct guide-dependent cleavage of foreign DNA (Willkomm et al., 2017; Willkomm et al., 2015; Zander et al., 2017). All together, these studies suggest that sRNAs are as widespread and abundant in the *Archaea* as in the *Bacteria* and *Eukarya*.

Target mRNA identification of sRNAs has proven to be difficult within the *Archaea* but a necessary task for uncovering sRNA functionality. RNA-seq in *M. mazei* cultures, grown under nitrogen starvation conditions, showed the differential expression of

a number of sRNAs in response to nitrogen availability, and allowed for the identification of the first *in vivo* target for archaeal intergenic sRNAs (Jager et al., 2012; Jager et al., 2009). The potential target for $sRNA_{162}$ is a bicistronic mRNA encoding for a transcription factor involved in regulating the switch between carbon sources and a protein of unknown function (Jager et al., 2012). In *Pyrobaculum*, 3 antisense sRNAs were found opposite a ferric uptake regulator, a triose-phosphate isomerase, and transcription factor B, supporting a potential role for archaeal antisense sRNA in the regulation of iron, transcription, and core metabolism (Bernick et al., 2012). sRNA deletion mutants can be used to identify potential biological functions and target genes. Deletion strains were successfully generated for *H. volcanii,* and phenotyping of the sRNAs deletion mutants revealed several severe growth defects under high temperatures, low salt concentrations, or specific carbon sources (Heyer et al., 2012; Straub et al., 2009). While these studies revealed that sRNAs likely play essential roles in the physiological response to environmental challenges in the *Archaea*, the functional roles and mechanisms of action of these important post-transcriptional regulators still remain unknown. Furthermore, no work has been done to investigate archaeal sRNAs in response to oxidative stress, a universal and frequent stressor in all domains of life that results in extensive cellular damage. In order to determine the impact of sRNAs during the oxidative stress response, we assessed the *H. volcanii* transcriptional landscape during no-challenge and oxidative stress conditions using comparative strand-specific small RNA-sequencing (sRNA-seq).

RESULTS

To identify globally small non-coding RNAs differentially expressed in response to oxidative stress in *H. volcanii*, we exposed 5 replicate cultures of *H. volcanii* to 2 mM H_2O_2, a dose that resulted in the survival of 80% of the cells **(Fig. 2-S1)**. RNA from these H_2O_2 treated cultures, and from no-challenge cultures (controls), were sequenced using a strand-specific size-selected (50-500 nt) sRNA library preparation essential for sRNA discovery.

Small non-coding RNA discovery in *H. volcanii* and normalized expression values. We obtained at total of 137 million sequence reads (41 Gb), across all replicates and conditions. Following quality control and reference-based read mapping, we intersected the mapped reads against the *H. volcanii* reference genome to discover sRNA transcripts that we classified as antisense (overlapping a gene and/or its regulatory elements on the opposite strand) (**Fig. 2-1A)** and intergenic (non-coding region between two genes) (**Fig. 2-1B).** We were unable to identify previously described *cis*-internal sRNAs (Babski et al., 2016) because, with our RNAseq strategies, those transcripts would be confounded with the corresponding gene transcripts. To further validate the sRNAs we identified, and reduce transcriptional noise in our data, we applied a rigorous two-pronged *in-silico* approach. First, we used a stringent threshold requiring that a sRNA (i) must be present in at least four out of five biological replicate libraries, and (ii) it must have a minimum expression of 40 transcripts per million (TPM, averaged with standard deviation between replicates) for antisense sRNAs and 14 TPM for intergenic sRNAs. Second, we used the genome browser IGV to inspect visually and confirm each sRNAs. These novel transcripts represented 25 to 30% of the total transcriptome and were 50 to 1,000 nucleotides (nt) in length (**Fig. 2-2**). Putative mRNA targets for antisense sRNAs were identified as the *cis*-

mRNA encoded on the opposite strand with a minimum overlap of 8 nucleotides as based on IGV confirmation and bacterial antisense sRNAs (Thomason & Storz, 2010). However, we found that the median overlap between sRNAs and putative *cis*-mRNA targets was 221 nucleotides (nt), with only a small number having a minimum overlap length of 8 nt (**Fig. 2-S2**). Analyzing the upstream regions of sRNAs enabled the discovery that 30% of sRNAs contained both a BRE and TATA-box with centroids at -38 and -29 nucleotides **(Fig. 2-S3)**. Using fewer conservative parameters (-3, +3 nucleotides) for BRE and TATA-box centroids resulted in 70% of sRNAs having transcriptional motifs.

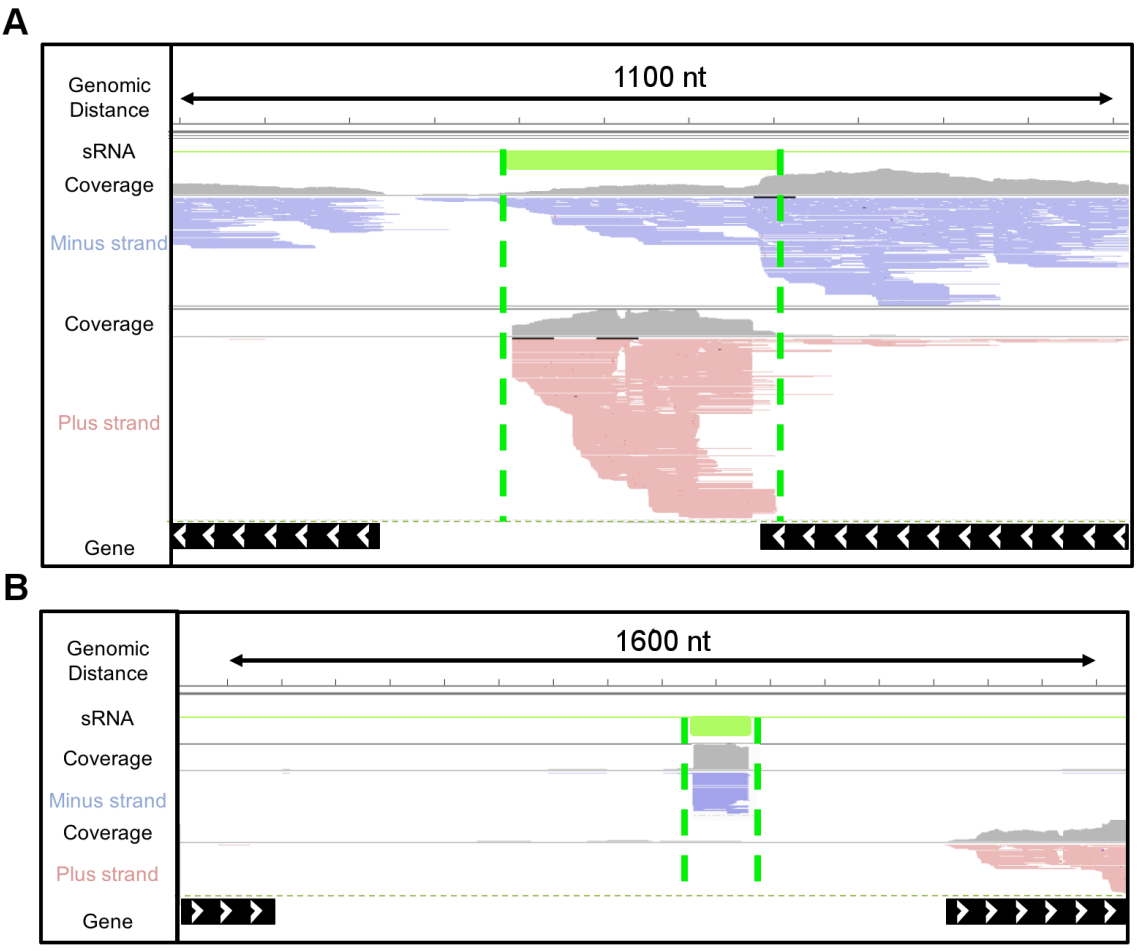

Figure 2-1: Genome viewer of (A) Antisense sRNAs (*cis*-acting) and (B) Intergenic sRNAs (*trans*-acting). Paired-end reads (100 bases) were mapped to the *H. volcanii* NCBI reference genome. Reference genes are marked as black lines with white arrows indicating their location on the plus strand (>) or minus strand (<). Reads marked in red are transcribed from the minus strand while blue reads are transcribed from the plus strand. Untranslated regions were predicted using Rockhopper2 (pink lines). Green lines mark discovered sRNAs. Coverage plots are in gray.

Normalized expression values in RNA-seq analyses are often reported as Reads or Fragments Per Kilobase of transcript per Million mapped reads (RPKM/FPKM). However, RPKM/FPKM have been shown to be inconsistent for comparison between samples due to variable transcript lengths. Another expression value, transcripts per million (TPM), was found preferable because it is independent of mean expressed transcript length and TPM normalization performs better in multiple library comparisons (Conesa et al., 2016; B. Li & Dewey, 2011; Wagner et al., 2012). To minimize transcript length bias (sRNAs are generally smaller in length) and to compare sRNA and mRNA expression levels, we chose to use TPM in our analysis.

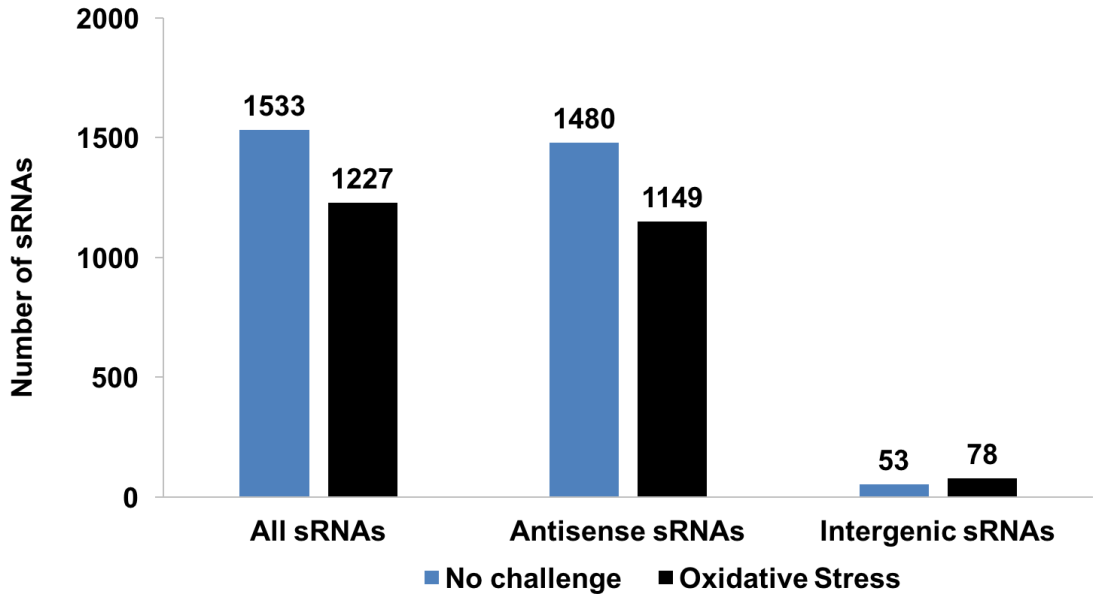

Figure 2-2: Number of sRNAs (total, antisense, and intergenic) discovered during no-challenge and H_2O_2 challenge conditions.

Non-coding sRNA characterization in *H. volcanii* during no-challenge conditions. *H. volcanii* grown under no-challenge conditions (42°C, complex media) expressed a total of 1,533 sRNAs after quality control (transcripts per million (TPM) ≥ 40) (**Fig. 2-2**), ranging from 49 to 1,000 nucleotides in size and with an average length of 373 nt. A majority of these sRNAs, 1,480 sRNAs (97%), were antisense to coding-regions (**Fig. 2-2**). The *H. volcanii* H53 auxotroph genome is 4 Mbp and contains 4,130 genes. The genome is comprised of a chromosome stably integrated with small chromosome pHV4, 2 small chromosomes (pHV1, pHV3), and has been cured of plasmid pHV2. A majority of sRNAs (68%) were encoded on the chromosome and integrated small chromosome pHV4 (18%). No sRNA encoded on plasmid pHV2 were found, as expected, while sRNAs were encoded on the remaining small chromosomes pHV1 (2%) and pHV3 (12%).

We compared mRNA expression to sRNA expression by constructing mRNA-seq libraries using the same RNA pool and library kit as the sRNA-seq libraries (omitting size-

selection) and calculating transcript expression as TPM (**Fig. 2-S4A**). Relative to mRNA expression levels (average: 312.1 ± 1079.1 TPM), the expression of the sRNAs was on average higher (average: 1107.9 ± 137.6 TPM). A comparison of expression levels between sRNAs and mRNAs further confirmed that a majority of sRNAs had higher expression levels than mRNAs (sRNA range: 14.1 to 905191.0 TPM, median: 108.5 TPM; mRNA range: 1.0 to 210162.0 TPM, median: 15.1 TPM). Overall, 75% of sRNAs had expression values less than or equal to 320 TPM, 15% had expression levels similar to that of highly expressed mRNAs, (500 to 10,000 TPM), and 16 sRNAs (sRNAs #1771 to #1786) had very robust expression levels with TPMs ranging from 10,000 to 60,000 TPM (**Fig. 2-S4A**). Lastly, one intergenic sRNA (STRG.2577.4), 111 nucleotides in size, exhibited expression levels higher than any mRNA with a TPM of 905,191. Transcript length did not correlate with expression levels, indicating that when we observed sRNAs with low expression levels, it was not an artifact of sequencing (i.e. longer transcripts receiving more read coverage thus skewing coverage based on length) **(Fig. 2-S5)**. We found that 4 of the 5 most highly expressed sRNAs (TPM > 30,00) were antisense to coding regions.

Putative mRNA targets of the most highly expressed asRNAs (≥ 10,000 TPM) included an IS4 Family Transposase (HVO_RS18385), ATP-cob(I)alamin adenosyltransferase (HVO_RS16235), transducer protein Htr36 (HVO_RS15355), pyridoxamine 5'-phosphate oxidase (HVO_RS07060), XerC/D integrase (HVO_RS01885), IS110 family transposase/pseudo region (HVO_RS07375), protein translocase TatA (HVO_RS09630), deoxyhypusine synthase (HVO_RS00895), sugar ABC transporter permease (HVO_RS17705), hydrolase (HVO_RS12225), peptidase (HVO_RS08770), RND transporter (HVO_RS14695), and IS110 family transposase

(HVO_RS02445). We do not report putative targets for intergenic sRNA because of the inherent difficulty in reliably predicting these targets due to unknown degrees of complementarity (i.e. gaps in hybridization between an intergenic sRNA and a mRNA).

Non-coding sRNAs in *H. volcanii* during oxidative stress conditions. *H. volcanii* under H_2O_2 –induced oxidative stress conditions expressed 1,227 sRNAs, a 20% decrease in the number of sRNAs compared to the no-challenge conditions (**Fig. 2-2**). Despite this decrease, a pattern of sRNA distribution similar to that of the no-challenge condition was observed; more than 94% of sRNAs were antisense and a majority (69%) were encoded on the main chromosome. A smaller average length of 337 nt was observed. Overall TPM expression of sRNAs during oxidative stress was similar to the no-challenge state, with a decrease in expression level for the single most highly expressed sRNA (STRG.2577.4 no-challenge: 905,191 TPM, STRG.2983.4 H_2O_2 : 810,120 TPM), which was an intergenic sRNA (**Fig. 2-S4B**). Putative targets for the most highly expressed antisense sRNAs (≥ 10,000 TPM) included both 16S rRNA genes (HVO_RS13015, HVO_RS18290; two sRNAs), an IS4 family transposase (HVO_RS18385), an M48 family peptidase (HVO_RS20365), an uracil-DNA glycosylase (HVO_RS09685), an agl cluster protein AglR (HVO_RS20160), hypothetical protein (HVO_RS20230), transducer protein Htr36 (HVO_RS15355), ISH3 family transposase/pseudo region (HVO_RS19935), IS4 family transposase (HVO_RS03870), (HVO_RS07060), and an ABC transporter ATP-binding protein (HVO_RS20010). Of these most highly expressed sRNAs, 3 antisense sRNAs (STRG.2050.2, STRG.3974.5, STRG.4700.1) targeted the same mRNAs and 3 intergenic sRNAs (STRG.3072.1, STRG.2702.1, STRG.2983.4) were expressed during both the no-challenge and oxidative stress conditions.

Regulatory effects and differential expression of sRNAs during oxidative stress. To investigate the regulatory effects sRNAs on their target mRNAs, we compared the expression levels (TPM) of all antisense sRNAs against the expression levels (TPM) of the *in silico* predicted putative *cis*-mRNA targets (**Fig. 2-3**). We found that a large population of asRNA-mRNA *cis*-pairs exhibited lower mRNA target expression compared to that of the sRNA, for both experimental conditions (**Fig. 2-3**). We conducted a pair-wise t-test between these *cis*-pairs on expression differences between sRNAs and putative *cis*-target mRNAs and found that 755 of sRNAs had significantly ($p < 0.05$) higher expression, potentially indicating a negative regulatory relationship between sRNAs and putative *cis*-mRNA targets (**Fig. 2-3**).

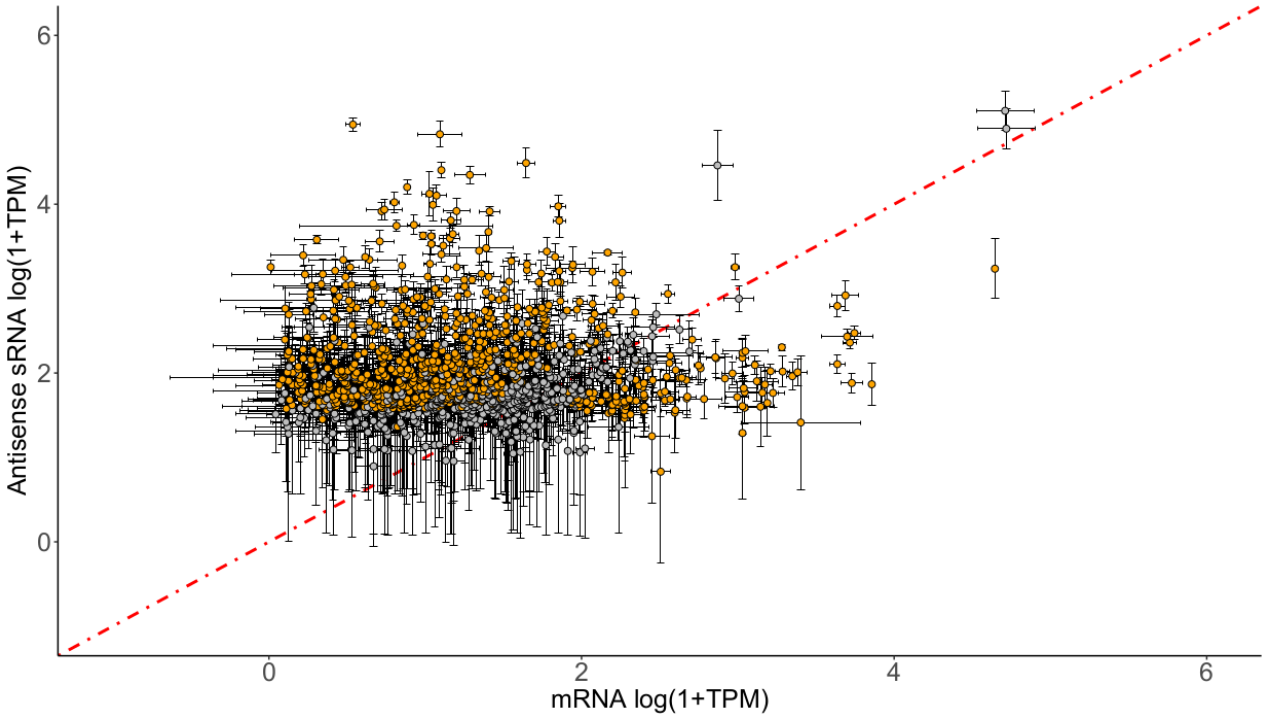

Figure 2-3: Transcript per Million (TPM) expression levels between sRNAs and their putative *cis*-mRNA targets during oxidative stress. Each point represents the TPM ratio between a sRNAs and its putative *cis*-mRNA target. TPM values are the averages among sRNA and mRNA replicates with error bars representing standard deviation among replicates. A pair-wise t-test was conducted between sRNA and putative *cis*-mRNA target replicates to infer significant difference in TPM expression between the *cis*-pairs. Orange points indicate a p-value < 0.05, and gray points indicate a p-value > 0.05. The red line represents no change in the ratio of sRNA-cis-mRNA expression (slope = 1).

To further investigate this negative regulatory relationship between sRNAs and putative mRNA targets we probed for differentially expressed sRNAs between the no-challenge and the oxidative stress conditions. Candidate sRNAs were considered significantly up- or down-regulated by oxidative stress using a False Discovery Rate (FDR) of less than 5%. Using this statistical framework, we identified a core set of differentially expressed sRNAs specific to oxidative stress (**Fig. 2-4A**). Both intergenic and antisense sRNAs were differentially expressed. Of the intergenic sRNAs, 48 were significantly differentially expressed (FDR < 0.05), with 23 up-regulated and 25 down-regulated **(Fig. 2-S6)**. Of these up-regulated intergenic sRNAs, 79% had greater than or equal to a fold-change of 2 ($\log_2$ fold-change = 1) increase in expression during oxidative stress, with the most up-regulated intergenic sRNA (STRG.277.2) having a 16 fold-change increase. On the other hand, a majority of down-regulated intergenic sRNAs had large fold-changes in expression (≥ 2 fold-change) and 5 exhibited very robust down-regulation (≤ -4 fold-change). A total of 605 antisense sRNAs were significantly (FDR < 0.05) differentially expressed. These sRNAs were either up-regulated (309) or down-regulated (296) during

oxidative stress, indicating two populations of antisense sRNAs (**Fig. 2-4A**). Fifty percent (302 sRNAs) of these differentially expressed sRNAs demonstrated a fold-change in expression of (±) 2 or greater; the most up-regulated sRNA had a fold-change of 15 and the most down-regulated sRNA had a fold-change of -9, indicating a role for these sRNAs in the cellular response to oxidative stress. More antisense sRNAs were up-regulated with a fold-change in expression of 4 or greater (28 sRNAs) compared to down-regulated (20 sRNAs). We then compared differential expression levels between all significantly up-regulated antisense sRNAs and their putative *cis*-mRNA targets and found that a population (133 sRNAs; 22%) of these up-regulated antisense sRNAs had putative mRNA targets that were down-regulated during oxidative stress (**Fig. 2-4B**). For example, during oxidative stress, 36 up-regulated antisense sRNAs targeted transposase mRNAs and 24 of these antisense sRNAs had putative *cis*-target transposase mRNAs down-regulated (**Fig. 2-3D**). While a smaller subset of significantly (FDR < 0.05) down-regulated antisense sRNAs had their putative *cis*-mRNA target up-regulated during oxidative stress (72; 11%), it also indicated a correlatively potential negative regulatory effect (**Fig. 2-4C**).

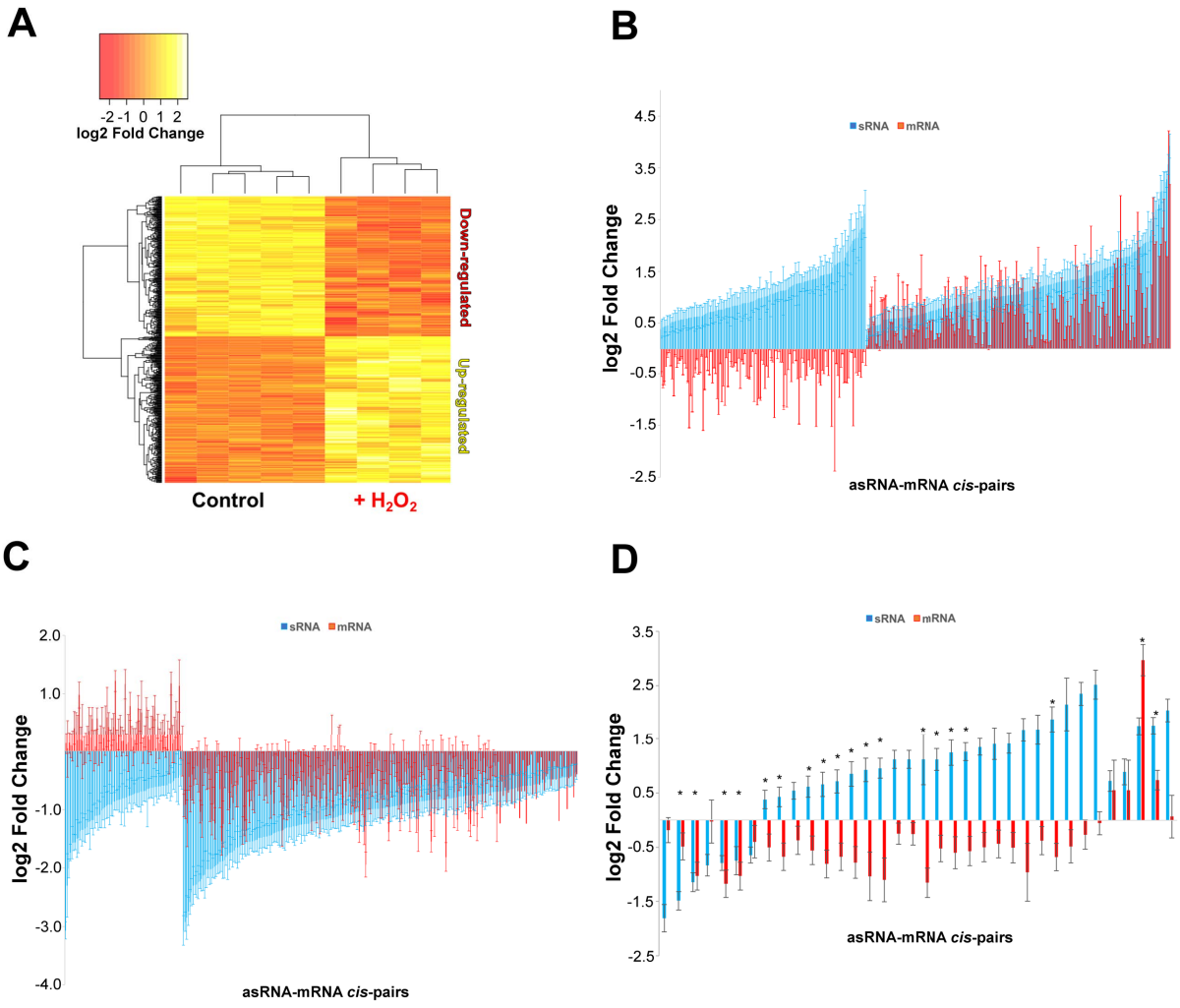

Figure 2-4: (A) Heatmap of $\log_2$ transformed fold-change of differentially expressed antisense sRNAs (asRNAs). (B) Differential expression fold-changes of up-regulated asRNAs and their putative *cis*-mRNA targets (averages and standard deviation error bars across replicates). (C) Differential expression fold-changes of down-regulated asRNAs and their putative *cis*-mRNA targets (averages and standard deviation error bars across replicates). (D) Differential expression fold-changes of all transposase-targeting asRNAs and their putative *cis*-transposase mRNA targets (averages and standard deviation error bars across replicates). * indicates that both the asRNA and the *cis*-target transposase mRNAs have significant differential expression based on FDR<0.05.

Oxidative-stress responsive antisense sRNAs were predicted to overlap both the 5' and 3' UTRs of mRNAs (**Fig. 2-5**). We found that 7% of antisense sRNAs overlapped at the 5' UTR and 26% overlapped at the 3' UTR. However, the majority of the antisense sRNAs (67%) were overlapped the coding sequence (CDS) of mRNAs rather than the UTRs, which has not been previously reported (**Fig. 2-5**). We calculated that, on average, the overlap between sRNAs and their putative *cis*-target mRNAs was 265 nt, the range was 8 to 992 nt, and the peak overlap was between 150 to 200 nt (**Fig. 2-S2**). Using Northern blots, we recapitulated the *in vivo* differential expression patterns of selected candidate sRNAs, further confirming transcript size and differential expression levels for oxidative stress sRNAs (STRG.3823.1, 4700.1, 8.6, 277.2, 3733.1, 2983.4, and 4213.4) (**Fig. 2-6A and 2-6B**). We also showed that the strandedness (the strand on which the sRNA was encoded) predicted by our sRNA-seq analysis was confirmed by our *in vivo* data using oligo probe northern blotting of 5' UTR, 3' UTR, CDS antisense sRNAs, and intergenic sRNAs (**Fig. 2-6B**).

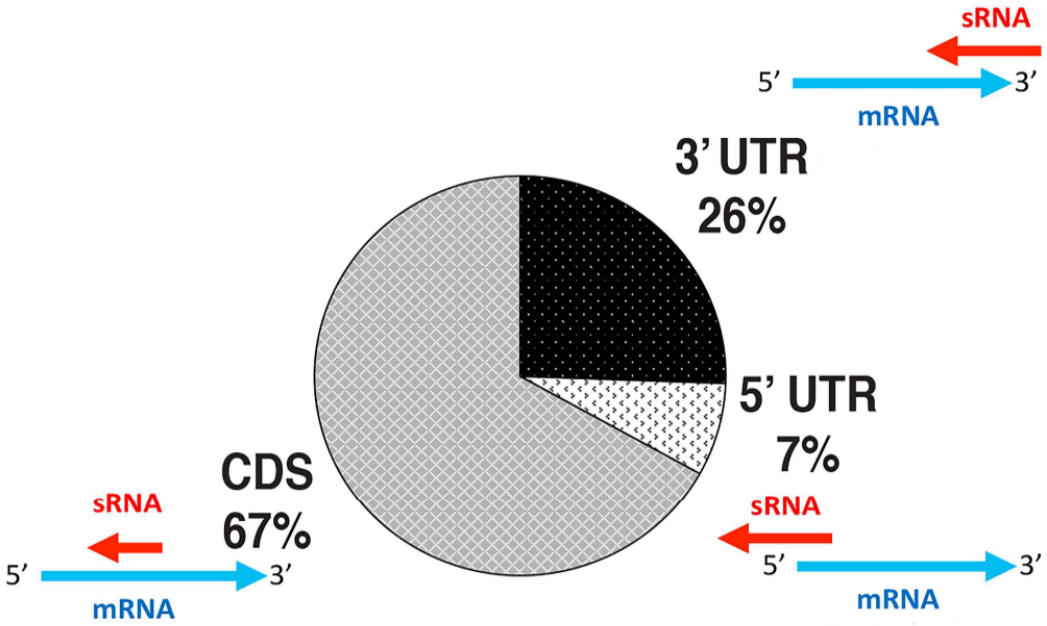

Figure 2-5: Distribution of binding regions for antisense sRNAs. UTR, untranslated region; CDS, coding sequence.

Target Enrichment of sRNAs. We identified *in silico* targets for the differentially expressed oxidative-stress responsive antisense sRNAs. Genes encoded by the putative target mRNAs were categorized by pathways using gene ontologies (GO) from the **D**atabase for **A**nnotation, **V**isualization and **I**ntegrated Discovery (DAVID). For sRNAs up-regulated during H_2O_2 stress, we found a functional enrichment of target genes encoding transposases, involved in chemotaxis methyl-receptor signaling and in transcriptional regulation (transcription factors) ($p < 0.05$). Genes, from many other pathways that were not enriched, were also the target of antisense sRNAs, including peptidase activity genes and serine and threonine biosynthesis genes (**Fig. 2-7A**). Twenty four of these sRNAs targeted transposase genes. Each transposase gene was down-regulated while their cognate sRNA was up-regulated, and the sRNA was predominantly located at the 5' UTR of its target. Most transposases belonged to the IS family of

transposases except for one DDE transposase. Three transcription factor families (IclR, ArcR, and Asn[C]) were also targeted by antisense sRNAs. A functional enrichment gene ontology analysis found that down-regulated sRNAs target genes were involved in membrane transport (ABC) transporters and biosynthesis of secondary metabolites, as well as targeting hydrolases (**Fig. 2-7B**). A significant proportion of enriched targets for both up- and down-regulated sRNAs were genes encoding hypothetical proteins.

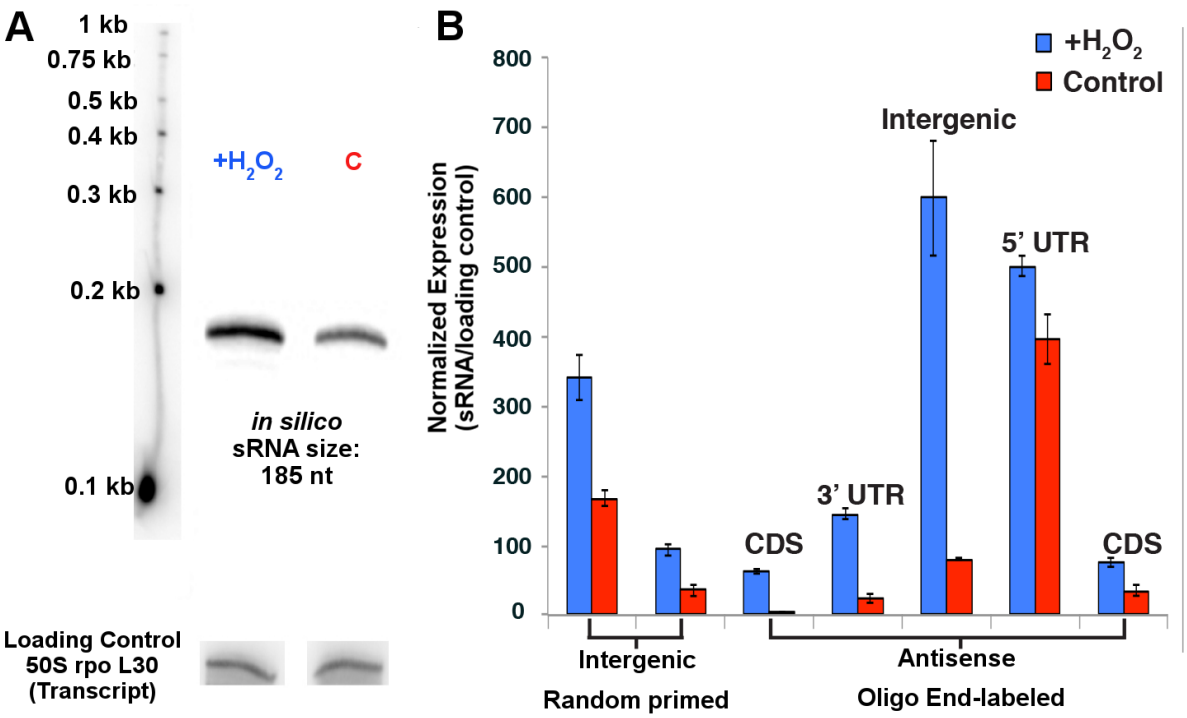

Figure 2-6: Validation of differentially expressed sRNAs by Northern blots. (A) Representative Northern blot confirming size and differential expression patterns of an intergenic sRNA during oxidative stress. (B) Quantification of Northern blots confirming the expression of intergenic sRNAs (random primed labeling) and strand-specificity of sRNAs (oligo labeling). All classes of sRNAs were confirmed: antisense (5' UTR, 3' UTR, CDS) and intergenic sRNAs.

mRNA transcriptional response to oxidative stress in *H. volcanii.* To determine the transcriptional landscape of mRNAs during oxidative stress, especially for mRNAs that were predicted targets of sRNAs, we sequenced rRNA-depleted mRNA-seq libraries in parallel with the previously described sRNA-seq libraries (derived from the same pool of total RNA). During H_2O_2-induced oxidative stress, a fourth of all genes (1,176) were significantly differentially expressed with a False Discover Rate less than 5%. Both catalase and superoxide dismutase, known ROS detoxification enzymes, were up-regulated at the mRNA level thus validating our experimental approach and characterizing *H. volcanii* response to oxidative stress at the transcriptional level (**Fig. 2-8**). A GO enrichment analysis (DAVID) was used to identify what pathways were enriched with differentially expressed genes during oxidative stress. The most enriched ($p<0.05$) up-regulated genes were involved in transcription, including various transcription factor families, all of the RNA polymerase subunit genes, and transcription initiation factors. Other enriched ($p<0.05$) up-regulated pathways were involved in iron-sulfur cluster assembly, DNA topological change (topoisomerase), proteasome, cell redox homeostasis, histidine metabolism, and 2-oxocarboxylic acid metabolism. The most up-regulated gene was a reactive intermediate/imine deaminase with a fold-change expression increase of 84. The most enriched ($p<0.05$) down-regulated genes were Tn5-like IS4 transposases. Other enriched ($p<0.05$) down-regulated pathways were pyrrolo-quinoline quinone (PQQ) proteins, tetrapyrrole methyltransferases, and ABC transporters. Only two genes had

down-regulation of less than fold-change -4, and these were an iron transporter and lysine 6-monooxygenase.

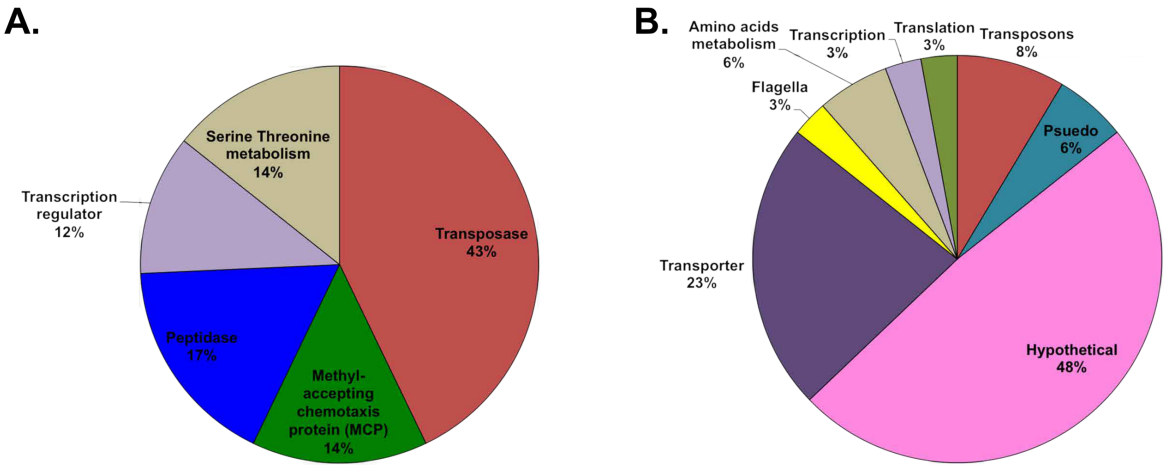

Figure 2-7: Gene ontology enrichment analysis identifying the functional classification of gene targets of sRNAs during oxidative stress. (A) Enriched target gene functions for up-regulated sRNAs. (B) Enriched target gene functions for down-regulated sRNAs.

DISCUSSION

Previous studies of sRNAs in *Archaea* revealed the abundance of sRNAs within the third domain of life and have been pivotal in establishing a working hypothesis on archaeal sRNA functionality. These studies have been limited to (1) microarray studies that do not allow *de novo* discovery of sRNAs, (2) differential RNA-seq approaches (dRNA-seq), which selects only for primary transcripts and does not provide length (nt) information, nor expression information (only coverage), and (3) individual sRNAs studies, which do not give a holistic view of the pathways being regulated within the cell. Using a custom strand-specific sRNA-seq library preparation and analysis pipeline, we have developed a method to perform high-throughput analysis of sRNA transcriptional landscape, expression, regulatory effects, and to identify regulated gene pathways in response to environmental

stressors within the *Archaea*. Through this study, we propose that sRNA-mediated transcriptional regulation is key in regulating stress responses to environmental challenges, such as oxidative stress, in the *Haloarchaea*. sRNAs have the potential to carry out large-scale regulation of genes involved in the oxidative stress response resulting in increased resistance to extreme environmental stressors.

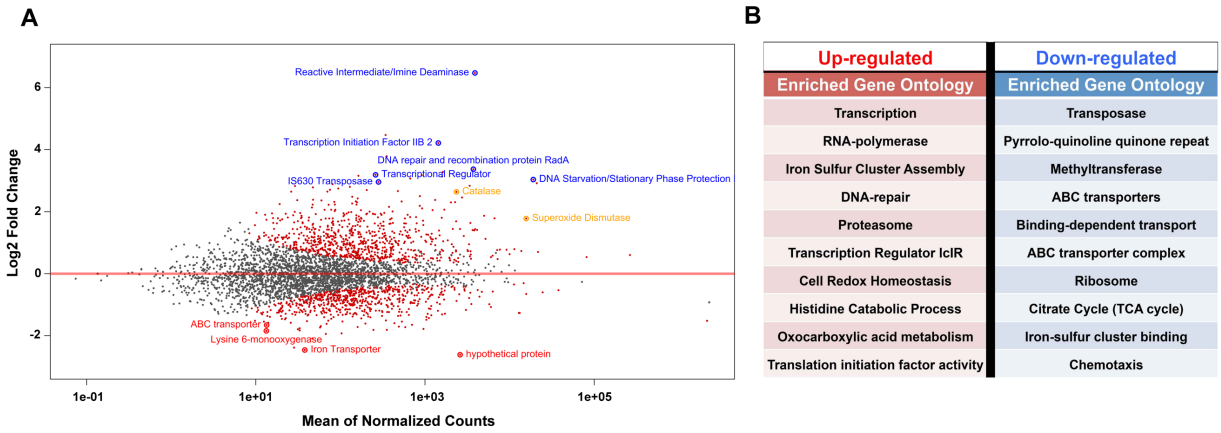

Figure 2-8: Distribution of differentially expressed genes during oxidative stress in *H. volcanii*. (A) MA-plot of differentially expressed genes; each point represents a gene. Significant (FDR < 5%) differentially expressed sRNAs are color-coded with up- (blue), down-regulated (red), and known oxidative stress response genes (yellow). (B) Gene function for the most up- and down-regulated sRNAs.

The discovery that sRNAs comprised nearly a third of the total transcriptome of *H. volcanii* and included basal transcriptional promoters, during both no-challenge and oxidative stress conditions, suggests that sRNAs have an important functional role under a variety of environmental conditions. We discovered thousands of sRNAs expressed in *H. volcanii* with the majority being antisense to genes, indicating that antisense transcription was ubiquitous within the cell. This is in stark contrast to most of the literature reporting that a majority of sRNAs discovered in *Archaea* were intergenic (Babski et al., 2014; Heyer

et al., 2012; Marchfelder et al., 2012; Straub et al., 2009). This discrepancy is likely due to previous studies being limited to microarray approaches. Indeed, a recent study using directional RNA-seq (dRNA-seq) to map all transcription start sites (TSS) in *H. volcanii* found thousands of novel transcript TSS with 1,244 of these TSS being antisense to mRNAs (Babski et al., 2016). Most of the TSS (75%) of the sRNAs we discovered in *H. volcanii* had the same TSS (+/- 5 nt) than those found in the dRNA-seq study by *Babski et al*, 2016 (Babski et al., 2016), further confirming our results. This underlines the importance of HTS studies, especially strand-specific RNA-seq such as our study, to discover the full extent of antisense sRNA expression in *Archaea*.

Our finding suggests that *cis*-acting sRNAs may play a larger role than *trans*-acting sRNAs within the cell but it should not be overlooked that the difficulty in finding *in silico* targets for intergenic sRNAs, because these sRNAs do not form 100% complementarity with their targets, might suggest that they have multiple mRNA targets. Antisense and intergenic sRNAs are broad classifications used in the archaeal small non-coding RNA field, but our data revealed that further classification can be done based on sRNA-mRNA binding characteristics (5’ UTR, 3’ UTR, CDS), differential expression, and regulatory effects. We found that only a small fraction of antisense sRNAs targeted the 5’ UTR of mRNAs, which is in concurrence with work demonstrating that most mRNAs in *H. volcanii* are mostly leaderless (lacking a 5' UTR). A majority of the 5’ UTR-binding sRNAs targeted transposons, providing further evidence that they may constitute their own class of sRNAs. Within this context, 3’ UTR-binding sRNAs should also be considered another class of sRNAs, resembling eukaryotic sRNAs, and likely acting on the degradation of their target transcript. The majority of the antisense sRNAs we identified in *H. volcanii* had 100%

complementarity within the CDS of their target mRNAs. This is the first report of such a finding in any domain of life and might constitute an attribute unique of archaeal sRNAs. We could not identify any 'seed' binding region for these CDS-binding sRNAs indicating that they likely have full occupancy upon the mRNA. It is worth noting that there were only a few instances (<20 total) where sRNAs overlapped more than the full length of the target mRNA (i.e. the asRNA is longer than the mRNA) or overlapped multiple *cis*-mRNA targets (ie. the asRNA overlaps the 3' UTR of one mRNA and the 5' UTR of an adjacent mRNA), which is novel in the Archaea.

Most *H. volcanii* sRNAs had a normalized expression value of 200 TPM or less indicating that sRNA transcripts are relatively abundant in the cell. Using a stringent thresholding approach resulted in a smaller number of more highly expressed sRNAs but avoided potentially reporting false positives with low TPM values or transcriptional noise. In comparison, most mRNAs within *H. volcanii* had TPM values of 20 or less. We found a population of antisense sRNAs with significantly higher expression compared to their putative *cis*-mRNA target, suggesting a potential negative regulatory role in sRNA-mRNA interactions (**Fig. 2-7, 2-8B**) (De Lay et al., 2013; He & Hannon, 2004; Storz et al., 2011). This trend extended to many asRNAs (both in no-challenge and oxidative stress conditions) down to sRNAs with 40 TPM in expression level. Stronger evidence for a negative regulatory effect lies with up-regulated sRNAs. A group of significantly up-regulated antisense sRNAs had target mRNAs that were significantly down-regulated during oxidative stress indicating these asRNAs may negatively regulate mRNA targets at the transcript level. Whether this negative regulation is occurring during transcription initiation/elongation or if these sRNAs are causing mRNA degradation is currently

unknown. All the sRNAs targeting transposons at the 5' UTR were up-regulated and the transposon mRNA were down-regulated (**Fig. 2-8D**), suggesting that these sRNAs might have a similar mechanistic function (Capy et al., 2000; Wheeler, 2013; Whitehead et al., 2006). If indeed these sRNAs are negatively regulating their target mRNAs in *H. volcanii*, we expected to find that down-regulated sRNAs have up-regulated target mRNAs. While some down-regulated sRNA targets exhibited this pattern, further supporting a potential negative regulation, many mRNA targets were also down-regulated (**Fig. 2-8C**). Alternative hypotheses, reflecting the complexity of transcriptional regulation in the Archaea, can be formed: (1) some of these sRNAs may have a positive regulatory effect, such as stabilizing target mRNAs and masking them from degradation, (2) *trans*-acting intergenic sRNAs might be targeting these mRNAs, negatively regulating them, and (3) some may have an unknown function (Prasse et al., 2017).

The most enriched negatively regulated sRNA targets were transposases, chemotaxis proteins, and transcription factors. It has been shown that transposons are opportunistic and, during stress conditions, can wreak havoc by hopping around in the genome causing double strand breaks, hence a need to be silenced (Capy et al., 2000; Wheeler, 2013; Whitehead et al., 2006). A functional enrichment of IS4 transposon genes being down-regulated during oxidative stress supports our observation that up-regulated sRNAs can potentially negatively regulate transposons (based on differential expression correlation between sRNAs and mRNAs) and suggests that transposon activity is tightly regulated during oxidative stress in *H. volcanii*. sRNA-mediated regulation of chemotaxis transducer proteins during oxidative stress suggests interesting implications in sensing ROS and motility. *H. volcanii* expresses a flagella analog named 'archaella', which is

organized into an operon and is regulated by a network of regulators called the archaellum regulatory network (arn) (identified in Crenarchaea) (Albers & Jarrell, 2015; Hoffmann et al., 2017). The regulation of these motility genes is still under investigation and so far is restricted to a few examples such as H_2/nitrogen limitation conditions in *M. janaschii and M. maripaludis* (Albers & Jarrell, 2015; Hendrickson et al., 2008; Mukhopadhyay et al., 2000; Xia et al., 2009). No direct transcriptional regulators of the archaellum have been identified in any euryarchaeota, but the deletion of archaellin genes, the presence of the H-domain set of type IV pillins, and agl proteins have been shown to affect the assembly of archaella in *H. volcanii* (Albers & Jarrell, 2015; Esquivel & Pohlschroder, 2014; Tripepi et al., 2010; Tripepi et al., 2012). Integral to how microorganisms maintain homeostasis in stressful and fluctuating environments are gene regulatory networks composed of interacting regulatory transcription factors and their target gene promoters (Darnell & Schmid, 2015). Our discovery that sRNAs are targeting transcription factors provides evidence that sRNAs are likely deeply interlaced within complex gene regulatory networks of *H. volcanii* and these sRNAs are key to maintaining homeostasis during environmental stress such as oxidative stress. Many mRNA-targets of differentially regulated sRNAs were hypothetical proteins, indicating that much remain to be elucidated in this organism.

Two single-stranded DNA binding proteins (RpaB and RpaC) were found to be required for increased survival of *H. volcanii* to ionizing radiation (a proxy for desiccation) and UV radiation, stressors that both cause oxidative stress (Skowyra & MacNeill, 2012; Stroud et al., 2012) (DiRuggiero lab, data unpublished). In *H. salinarum*, Rpa operons were up-regulated during ionizing radiation as well and contributed to resistance (DeVeaux et al., 2007; McCready et al., 2005). In conjunction to previous findings, we observed that

two of the most up-regulated genes during H_2O_2 oxidative stress were RpaB1 (HVO_RS10725: 23 fold-change) and RpaB2 (HVO_RS06105: 8 fold-change) confirming their role in oxidative stress resistance in *H. volcanii* and likely other haloarchaea. One gene, a reactive intermediate/imine deaminase RidA-homolog (HVO_RS12485), was up-regulated orders of magnitude more than any other gene. The encoded protein is known to be involved in synthesis of branched-chain amino acids by speeding up the IlvA-catalyzed deamination of threonine into 2-ketobutyrate (Lambrecht et al., 2012; Müller et al., 2014). Previous work has shown that in the presence of reactive chlorine species (RCS), such as HOCl, imine deaminase seemed to inhibit IlvA activity suggesting that imine deaminase may have a different function in the presence of RCS (Dahl et al., 2015; Müller et al., 2014). Further studies found that imine deaminase can sense RCS and in doing so becomes a chaperone that prevents protein aggregation (Müller et al., 2014). Reactive oxygen species in hypersaline environments produce RCS (Stutz et al., 2002). In addition, ROS causes extensive, irreversible protein damage such as carbonylation, which in turn causes protein aggregation (Nyström, 2005; Suzuki et al., 2010). This reactive intermediate/imine deaminase is the most up-regulated protein-encoding gene, suggesting that it may be playing a similar chaperon role to prevent protein aggregation, either sensing ROS or RCS produced by H_2O_2 (Dahl et al., 2015; Müller et al., 2014).

This is the first study to report on the transcriptional response of *H. volcanii* to oxidative stress and, while we found similar responses to H_2O_2 exposure than previously reported for *H. salinarum*.(Kaur et al., 2010), further validating our work and providing evidence that *Haloarchaea* have evolved similar strategies to survive their environmental stresses, we also found responses that were unique to *H. volcanii.* Similarities to *H.*

salinarum include the up-regulation of ROS scavenging proteins (catalase, superoxide dismutase), iron sulfur assembly proteins (SufB, SufD), proteasome genes, indicating high protein turn-over, and many DNA-repair genes (Kaur et al., 2010). Most of the down-regulated genes were involved with metabolism, such as sugar/phosphate/peptide ABC transporters, electron carriers (halocyanin), and TCA cycle enzymes, possibly to halt growth until damage is repaired (Kaur et al., 2010; Scharf & Engelhard, 1993). The most down regulated gene was an iron ABC transporter, most likely to limit further production of ROS via Fenton reactions (Kaur et al., 2010). Of unique responses to oxidative stress in *H. volcanii,* we found that all of the RNA polymerase subunits and transcription elongation factors, and seven basal transcription initiation factors (HVO_RS12755: 18.4; HVO_RS01380: 8.6; HVO_RS09745: 6.5; HVO_RS01840: 2.1; HVO_RS05475: 1.74; HVO_RS11835: 1.5; HVO_RS05475: 1.3 fold-change) were significantly up-regulated in response to oxidative stress. The increase in sRNAs during oxidative stress could be attributed to this increase in transcription machinery. The majority of the 30S and 50S ribosomal subunits were down regulated, in contrast to *H. salinarum*. The up-regulation of histidine biosynthesis and catabolism into glutamate, and 2-Oxocarboxylic acid metabolism were unknown to be involved in the oxidative stress response, which further demonstrates there are still more mechanisms to uncover for oxidative stress resistance. RosR was identified as a global transcriptional regulator in *H. salinarum* and it strongly up-regulated during oxidative stress (Sharma et al., 2012). RosR demonstrated no differential expression to oxidative stress in *H. volcanii* indicating that it may play another role in this organism. Cell cycle genes (parA, cdc6) involved in chromosome

segregation(Lindas & Bernander, 2013) were down regulated, further suggesting that division is being arrested (halting growth) in order to repair damage.

In this study, we showed for the first time that small non-coding RNAs are specifically associated with the oxidative stress response in Archaea. During oxidative stress, antisense sRNAs were prevalently transcribed, comprising nearly 30% of the transcriptome of *H. volcanii*, and many up-regulated antisense sRNAs imparted a correlative negative regulatory effect on target mRNAs. These results support the hypothesis that antisense sRNAs in Archaea behave similarly to *cis*-acting bacterial sRNAs and eukaryotic siRNAs, which negatively regulate mRNAs by sharing extensive complementarity and facilitating RNA degradation (Mack, 2007; Waters & Storz, 2009). The precise mechanism(s) of sRNA-mRNA mediated regulation remains to be elucidated and in particular whether proteins are required to complex with sRNAs in order to mediate gene regulation such as in Bacteria (Hfq) and Eukarya (Ago). We also identified several classes of antisense sRNAs, based on their mRNA-binding patterns (3' UTR, 5' UTR, and CDS), and showed that CDS-targeting of mRNAs was the predominant mode of action for sRNA hybridization. Mechanistic differences between these classes of sRNA still need to be investigated as well as the regulatory roles of sRNAs in Archaea and their functional importance in adaption to extreme environments.

MATERIAL AND METHODS

Culture growth conditions. *H. volcanii* auxotrophic strain H53 (*Δpyre2, ΔtrpA*) was used for all experiments. Culturing in liquid and solid media was done in rich medium

(Hv-YPC), at 42°C and with shaking at 220 rpm (Amerix Gyromax 737).(Halohandbook) Uracil and tryptophan were added to a final concentration of 50 µg/mL, each.

Oxidative stress exposure. We exposed 5 biological replicates of *H. volcanii* strain H53 liquid cultures to the oxidative stress agent H_2O_2. Initially, cultures were grown in 80 mL of Hv-YPC under optimal conditions to an OD of 0.4 (mid exponential phase). To ensure homogeneity, each replicate was subsequently split into two 40 mL cultures, one used for the no-challenge (control) condition and the other for the oxidative stress condition. For the latter condition, 2 mM H_2O_2 (80% survival rate, previously determined) was directly added to the cultures followed by an hour incubation at 42 °C with shaking at 220 rpm. Cultures were then rapidly cooled down, centrifuged at 5,000 x g for 5 minutes and the pellets resuspended in 18% sea water. The cell suspensions were then transferred to a 1 mL tube and centrifuged at 6,000 x g for 3 minutes, the pellets were flash frozen and stored at -80 °C until ready for RNA extraction. Control no-challenge culture replicates were processed in the same manner without the addition of H_2O_2.

Oxidative stress survival curves. Assessment of survival in *H. volcanii* under oxidative stress conditions was done using microdilution plating as described before.(C. K. Robinson et al., 2011) Counts were averaged and standard deviation calculated between replicates. Survival was calculated as the number of viable cells following H_2O_2 treatment divided by the number of viable untreated cells and graphed with standard error bars.

RNA extraction. Total RNA was extracted using the Zymo Quick-RNA Miniprep kit with the following modifications: after addition of RNA lysis buffer to the frozen cell pellets, cells were processed with a 23 G needle and syringe to insure complete cell lysis. *H. volcanii* liquid culture is slimy and viscous thus to increase cellular lysis a 23 G needle

and syringe was used to break down the cell pellet. Total RNA was then extracted following the standard kit protocol.

Small RNA-sequencing library preparation (sRNA-seq). Total RNA, for each biological replicate and condition, was size-selected using denaturing polyacrylamide gel electrophoresis. 20 µg of total RNA was loaded onto a 7% denaturing urea polyacrylamide gel (SequaGel, National Diagnostics) in 0.5 x TBE buffer and ran at constant power of 30 W until bromophenol blue bands reached the bottom of the gel. The gel was stained with SYBR Gold, visualized on a blue light box, and bands in the 50-500 nucleotide range, as indicated by the RNA Century Marker plus ladder (ThermoFisher), were excised. Small RNAs (sRNA) were eluted by rotating overnight in 1.2 mL 0.3 M NaCl, ethanol precipitated, and DNase I (NEB) treated (37 °C for 2 hours) as previously described.(Zhang, Theurkauf, Weng, & Zamore, 2012) Strand-specific libraries were prepared using the SMART-seq Ultralow RNA input kit (Takara), insert sizes checked with the Bioanalyzer RNA pico kit (Agilent), and paired-end sequencing (2 x 150 bp) was carried out on the Illumina HiSeq 2500 platform at the Johns Hopkins University Genetic Resources Core Facility (GRCF).

Messenger RNA-sequencing library preparation (mRNA-seq). Individual mRNA-seq libraries were made from the same RNA pools as the sRNA-seq libraries above. Total RNA was rRNA-depleted using the Ribo-zero Bacteria kit (Illumina) and sequenced using the sRNA-seq library preparation method described above but omitting the size-selection by denaturing urea polyacrylamide gel.

sRNA- and RNA-seq read quality control and reference-based read mapping. Assessment of the quality of each sequencing library read was determined using fastqc.

The program trim galore was used with base settings to trim adapter sequences from reads and to filter out low phred score reads (<20). Short length reads were preserved. Reads from each replicate were aggregated together per condition to get a set of consensus sRNAs and were mapped against the *H, volcanii* NCBI refseq genome (taxid 2246; 1 chromosome, 4 plasmids) using the hisat2 aligner with strand-specific options turned on and splice aware options turned off, paired-end mode.(Kim, Langmead, & Salzberg, 2015)

sRNA- and RNA-seq transcriptome assembly. The reference-based alignments were assembled into transcriptomes using the program stringtie in order to build full-length transcripts, calculate coverage and expression values (TPM). The assembly was guided by a gene annotation file from the *H. volcanii DS2* (NCBI refseq taxid 2246) genome to build transcripts and annotate them either as a gene or novel transcript.(Pertea et al., 2015) A minimum distance between reads for transcript assembly was specified at 30 nucleotides. gffcompare under default options was used to compare the assembled transcriptomes against the gene annotation file from *H. volcanii DS2* (NCBI refseq taxid 2246) to annotate transcripts as genes or non-coding RNA (antisense or intergenic).(gffcompare; Pertea et al., 2016) In house python scripts were used to bin transcripts that were annotated as genes, transcripts annotated as antisense (classified as non-coding region opposite from a coding region), transcripts annotated as intergenic (classified as non-coding region between two coding regions), and subsequently binned antisense sRNAs as 3' UTR, 5' UTR, or CDS overlapping.

sRNA- and RNA-seq differential expression analysis. We used a read count-based differential expression analysis to identify differentially expressed sRNAs during oxidative stress. The program featureCounts was used to rapidly count reads that map to the

assembled sRNA transcripts (described above).(Liao et al., 2014) featureCounts was run with strand-specific options on, paired-end mode on, multi-mapping off.(Liao et al., 2014) The read counts were then used in the R differential expression software package DESeq2.(Love et al., 2014) Briefly, read counts were converted into a data matrix and normalized by sequencing depth and geometric mean. Differential expression was calculated by finding the difference in read counts between the no-challenge state (control) normalized read counts from the oxidative stress normalized read counts.(Love et al., 2014) The differentially expressed sRNAs were filtered based on the statistical parameter of False Discovery Rate (FDR) and those that were equal to or under a FDR of 5% were classified as true differentially expressed sRNAs.

in silico validation of sRNAs. Differentially expressed sRNAs were validated by two *in silico* methods: 1) Visualization of transcripts, and 2) open reading frame protein homology search. In the first method, transcriptomes for each replicate and condition were visualized on the Integrated Genome Viewer (IGV) against the *H, volcanii* (NCBI refseq taxid 2246) genome and annotation.(J. T. Robinson et al., 2011) The sRNA transcript coordinates were used to locate putative sRNAs and if it was found within an operon it was eliminated from further analysis. In the second method, blastx (default parameters) was used to search for protein and domain homology for each sRNA and those that had significant homology with known proteins or domains were eliminated from further analysis.(Altschul et al., 1990)

Regulatory element motif identification of sRNAs. 100 nucleotides upstream and downstream from the sRNA transcript start and stop coordinates were extracted using in house python scripts. These regions were searched for transcription motifs (BRE, TATA-

box) using both multiple sequence alignments and visualization with WebLogo (default parameters) and motif searching with MEME and CentriMo (default parameters).(Bailey et al., 2009; Crooks et al., 2004)

In vivo validation of sRNAs by Northern Blot analysis. 20 µg of total RNA and P^{32} ATP end-labeled Century+ RNA markers were loaded onto 5% denaturing urea polyacrylamide gels (SequaGel, National Diagnostics) and run at 30 watts for 1.5 hours to ensure well-spaced gel migration from 50 to 1,000 nucleotides (nt). Gels were transferred onto Ultra-hyb Nylon membranes and hybridized with 2 types of probes. For lowly expressed sRNAs, we probed with [γ-P^{32}]dATP randomly primed amplicons generated with custom primers based on sRNA transcript genomic coordinates as determined by the sRNA-seq *in silico* analysis. Probe primers were at a minimum 10 nt inwards from the predicted genomic coordinates (start and stop) to ensure accurate transcript detection. Hybridizations were done at 65°C. To determine strandedness of sRNAs, we used [α-P^{32}]dATP end-labeled oligo probes (20-23 nt) that were antisense to sRNAs. Hybridizations were at 42°C. The rpl30 protein (HVO_RS16975) transcript was used as a loading control for differential expression calculation because it was not differentially expressed under oxidative stress in this RNA-seq dataset. Differential expression was calculated using ImageJ.

Gene Ontology (GO) enrichment analysis of mRNA-targets. NCBI gene names for all mRNA-targets of antisense sRNAs were uploaded into **D**atabase for **A**nnotation, **V**isualization and **I**ntegrated Discovery (DAVID) to determine the pathways and gene ontologies targeted by sRNAs.

RNA-seq data. All raw read and processed data from these experiments are available at NCBI under BioProject PRJNA407425. Illumina raw sequence data (.fastq) for each replicate and condition are deposited in NCBI Sequence Read Archive with accession number SRP117726.

CONTRIBUTIONS

DRG: Conceptualization, Investigation, Methodology, Project administration, Writing – original draft, Writing – review & editing

JDR: Conceptualization, Funding acquisition, Project administration, Supervision, Validation, Writing – review & editing

ACKNOWLEDGEMENTS

This work was supported by grant FA9950-14-1-0118 from the AFOSR.

We thank Madeline Cassani, Dr. Zhao Zhang, and German Uritskiy for advice and guidance on sRNA-seq library preparation, Evan Hass and Dr. Vidya Balagopal for advice on northern blotting, Dr. Jacques Ravel, Mike Humphrys, and David Mohr for sequencing efforts and technical advice, and Dr. John Kim and Dr. Sarah Woodson for helpful discussions.

SUPPLEMENTAL FIGURES

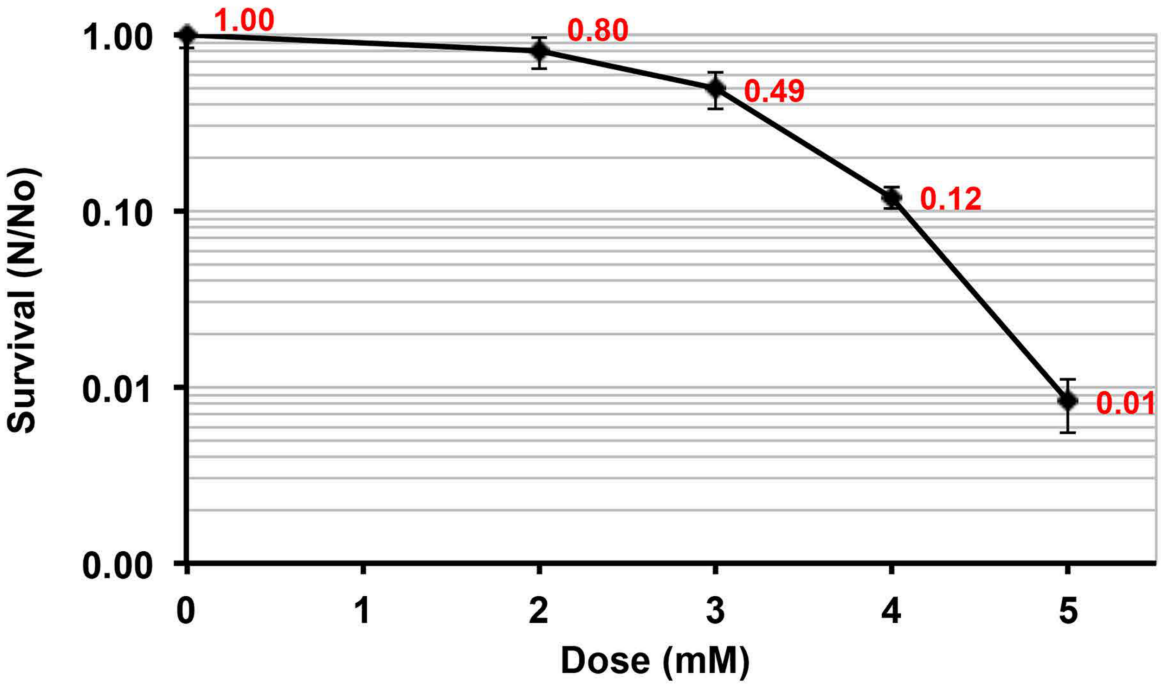

Figure 2-S1: Survival curve of *H. volcanii* to hydrogen peroxide exposure. Survival is represented by colony forming units.

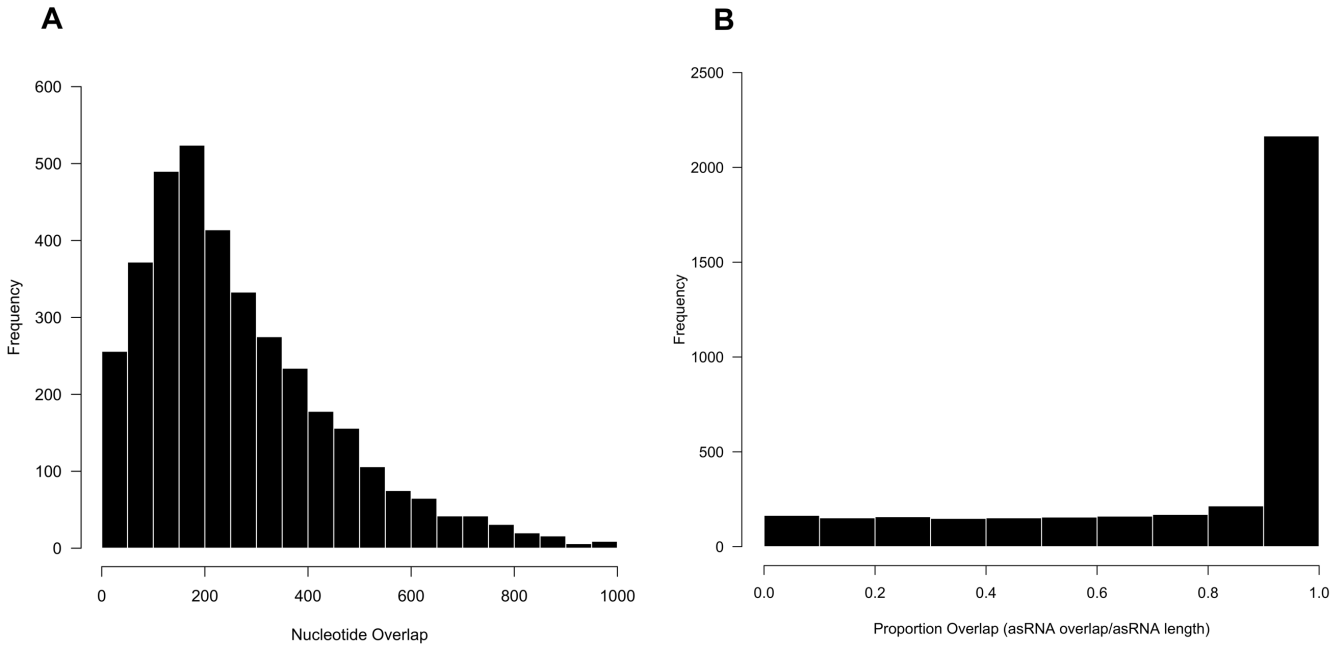

Figure 2-S2: Characterizing asRNAs and their overlaps with their corresponding genes. (A) Histogram of nucleotide overlap between asRNAs and their corresponding genes. (B) Histogram of the proportion of the asRNA that overlaps with a gene.

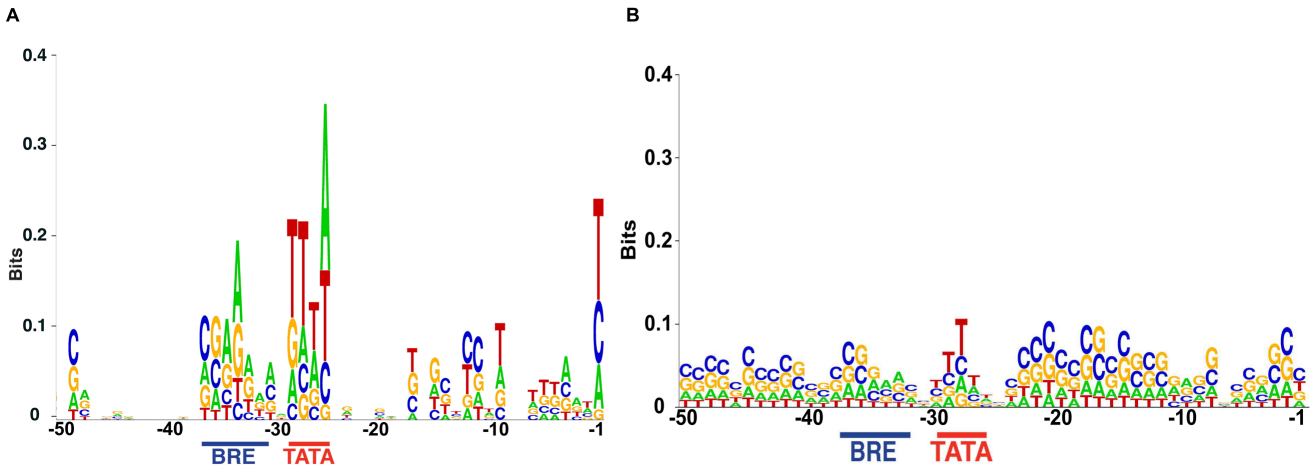

Figure 2-S3: Weblogos of 50nt upstream of the transcription start site for (A) identified sRNAs, both its and asRNAs, and (B) for mRNAs.

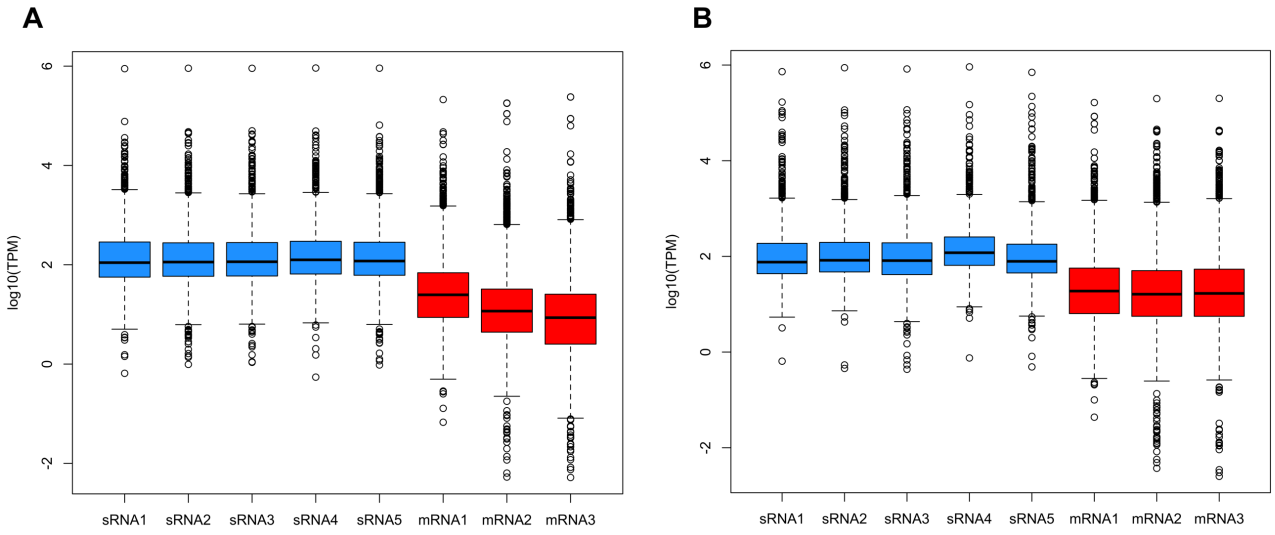

Figure 2-S4: Distribution of expression levels (transcript per million; TPM) of sRNAs and mRNAs detected in RNA-seq libraries under (A) no challenge conditions and (B) oxidative stress conditions.

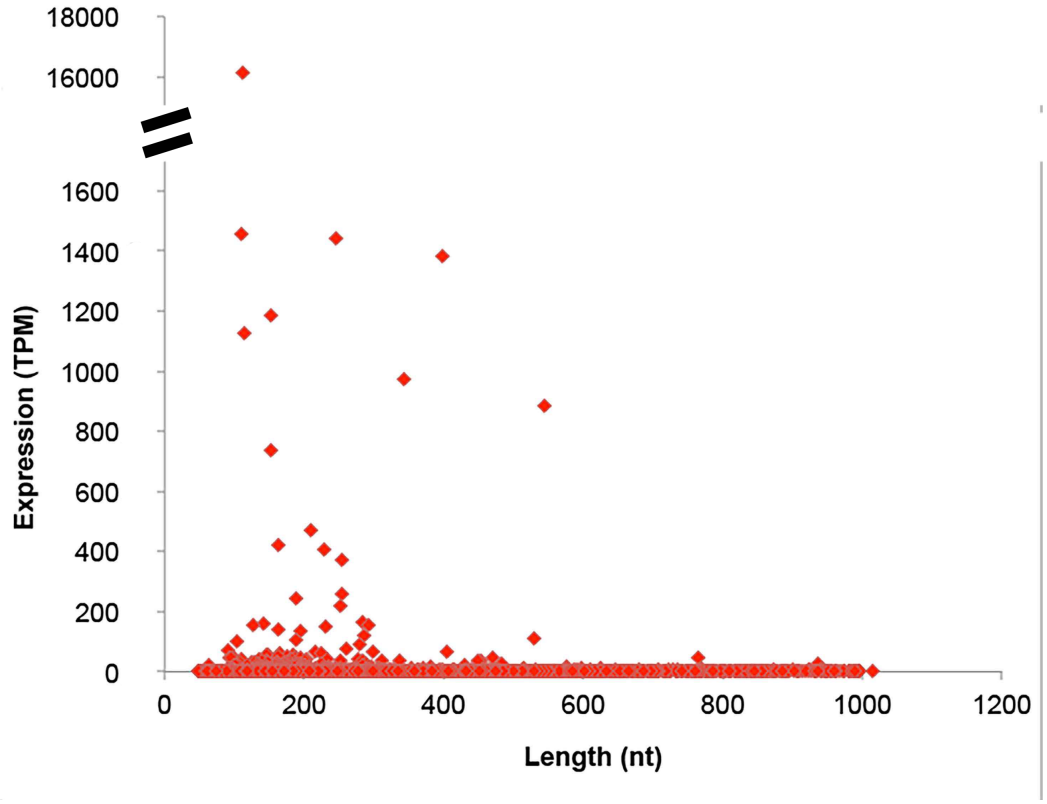

Figure 2-S5: Expression levels (TPM) of sRNAs relative to their length (nt).

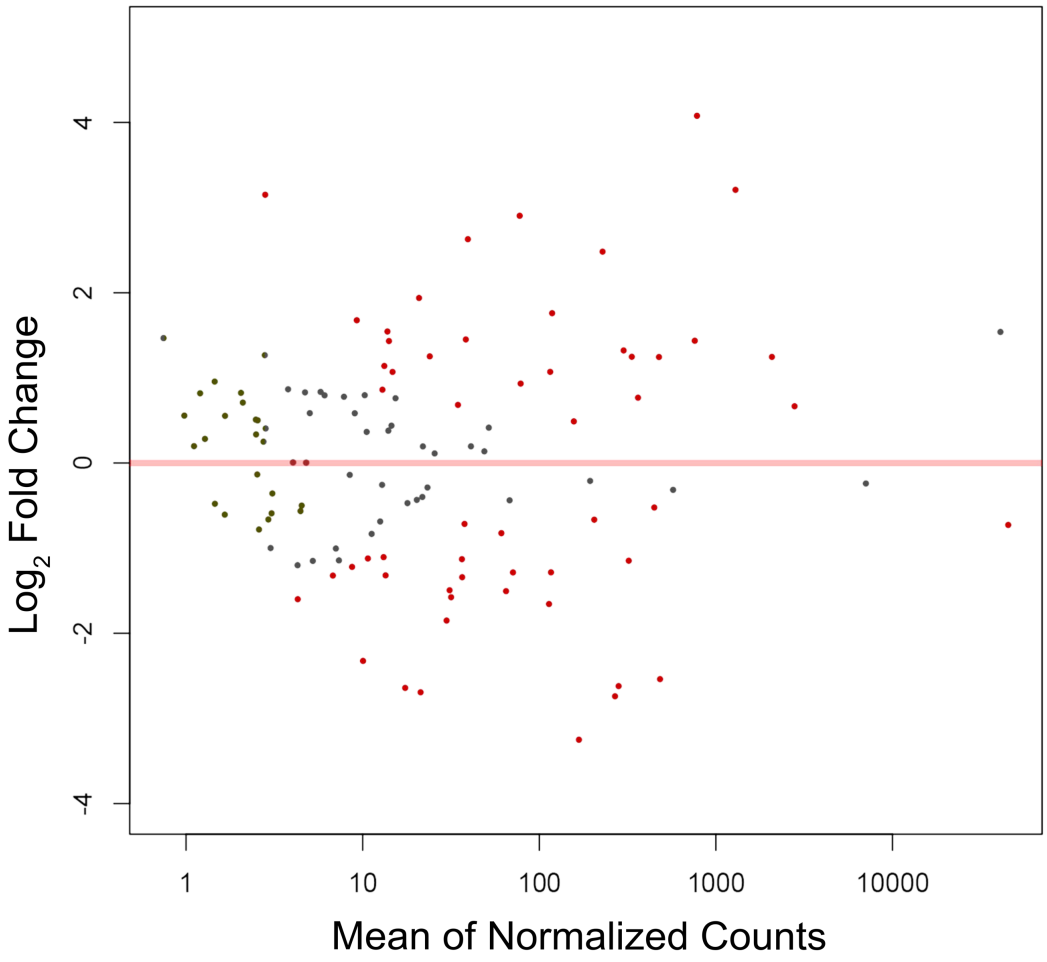

Figure 2-S6: MA-plot of differentially expressed intergenic sRNAs (itsRNAs). Each point represents a sRNA and red points are significant differential expression (False Discovery Rate < 0.05).

Chapter 3

Ribosome Profiling in Archaea Reveals Leaderless Translation, Novel Translational Initiation Sites, and Ribosome Pausing at Single Codon Resolution

This work has been submitted to Nucleic Acids Research January 2020 – Gelsinger DR, Dallon E, Reddy R, Mohammad F, Buskirk A, & DiRuggiero J, Nucleic Acids & Research January 2020

ABSTRACT

High-throughput methods, such as ribosome profiling, have revealed the complexity of translation regulation in Bacteria and Eukarya with large-scale effects on cellular functions. In contrast, the translational landscape in Archaea remains mostly unexplored. Here, we developed ribosome profiling in a model archaeon, *Haloferax volcanii*, elucidating, for the first time, the translational landscape of a representative of the third domain of life. We determined the ribosome footprint of *H. volcanii* to be comparable in size to that of the Eukarya. We linked footprint lengths to initiating and elongating states of the ribosome on leadered transcripts, operons, and on leaderless transcripts, the latter representing 70% of *H. volcanii* transcriptome. We manipulated ribosome activity with translation inhibitors to reveal ribosome pausing at specific codons. Lastly, we found that the drug harringtonine arrested ribosomes at initiation sites in this archaeon. This drug treatment allowed us to confirm known translation initiation sites and also reveal putative

novel initiation sites in intergenic regions, on non-coding RNAs, and within genes. We further demonstrated that some alternative translation start sites were regulated during oxidative stress. Ribosome profiling revealed an uncharacterized complexity of translation in this archaeon with bacteria-like, eukarya-like, and potentially novel translation mechanisms. These mechanisms are likely to be functionally essential and to contribute to an expanded proteome with regulatory roles in gene expression.

INTRODUCTION

Archaea stand at the crossroad between the other two domains of life, using a mosaic of molecular features from both Bacteria and Eukarya along with unique characteristics. For example, Archaea have a eukaryotic-like transcription apparatus but bacterial-like transcription regulation and coupled transcription-translation (Ferry & Kastead, 2007). Moreover, Archaea have polycistronic mRNAs like Bacteria, implying that archaeal ribosomes can perform repeated cycles of initiation on the same mRNA (Brenneis et al., 2007; French et al., 2007; Santangelo et al., 2008). Archaeal ribosome subunits (30S/50S) and rRNA genes (16S/23S) are closer in size to their bacterial counterparts rather than to Eukarya (Lecompte et al., 2002). Even with these similarities, Archaea are still evolutionarily closer to Eukarya, and recently it has been proposed that the newly discovered Asgard archaea may be sister taxa to the Eukarya (Eme et al., 2017; Imachi et al., 2019; Spang et al., 2015; Williams et al., 2013). This close evolutionary relationship between the two domains is based on similarities of central dogma processes (replication, transcription, and translation) and an expected complexity in the Archaea (Bell & Jackson, 1998; Grabowski & Kelman, 2003). We thus stand to discover novel mechanisms and processes by studying the molecular biology of Archaea.

Translation is essential for life and has been proposed to be the first subsystem to "crystallize" during the transition of pre-cellular to cellular life as we know it (Woese, 1998). Despite evolutionary conservation, the translation apparatus, the ribosomes, and their accessory proteins have diverged in the three domains of life (Bell & Jackson, 1998; Benelli & Londei, 2011; Clouet-d'Orval, 2017). The bacterial translation system is highly streamlined, while Archaea and Eukarya have evolved an expansion of initiation and recycling factors that causes a high degree of complexity in the translation cycle and

suggests a need for extensive regulation of this process (Bell & Jackson, 1998; Clouet-d'Orval, 2017; Dennis, 1997; Hinnebusch & Lorsch, 2012).

Even after decades of research, our understanding of translation in Archaea remains limited compared to the other domains of life. There is also a puzzling undeniable link between the archaeal and eukaryotic translation apparatus that remains to be addressed. The lack of knowledge in archaeal translation mechanisms is primarily due to the difficulty in cultivating these organisms because many archaeal model organisms are extremophiles, making it difficult to use classical biochemical and molecular methods. While substantial work has been done to characterize archaeal mechanisms of translation initiation *in vitro* (Akulich et al., 2016; Benelli & Londei, 2011; Lo Gullo et al., 2019; Schmitt et al., 2019), these studies were focused on a few select genes as opposed to a global approach. As a result, it remains difficult to predict a unified view of translation processes and their regulation across all mRNAs in an archaeon where many different transcriptional units co-exist (i.e. leadered, leaderless, and operonic transcripts). There is, therefore, a clear need for a high-resolution, genome-wide view of translation in the Archaea.

To address the current knowledge gap in the mechanisms and regulation of translation in Archaea, we developed a simple, robust, and reproducible protocol for ribosome profiling in the model halophilic archaeon *Haloferax volcanii*. *H. volcanii*, a member of the *Halobacteria*, is adapted to high salt and uses a "salt-in" strategy (accumulation of 2-4 M KCl within the cell) to balance the osmotic pressure from its high salt environment (Oren, 2013). We used ribosome profiling to gain a high-resolution view of various aspects of archaeal translation, including the ribosome footprint size of

archaea, leaderless initiation, and global ribosomal pausing. Going further, we used harringtonine to stall ribosomes at initiation codons and identified novel and alternative translation initiation sites genome-wide. We lastly investigated the translational control of alternative open reading frames during oxidative stress to characterize the regulatory potential of a stress resistant extremophile. This work describes novel insights into archaeal translation processes and provides an experimental paradigm for the *in vivo* study of translation at high salt. It also provides a framework for the adaptation of this technique to other Archaea.

RESULTS

Development of a ribosome profiling method for Haloferax volcanii

Ribosome profiling is the deep sequencing of ribosome-protected mRNA footprints (RPFs) after nuclease digestion. Developing this method for *H. volcanii* provided, for the first time, a genome-wide survey of translation at a level of resolution currently unknown in the Archaea. The method has four critical steps (**Fig. 3-1A**): (1) inhibit translation so that ribosomes are arrested as samples are processed, (2) digest unprotected mRNA with nucleases, (3) purify 70S monosomes from sucrose gradients, and (4) convert RPFs to dsDNA for deep sequencing (Mohammad et al., 2019). Ensuring that these steps are reproducible and robust in *H. volcanii* was particularly challenging due to its extremely high intracellular salt concentration (2-4 M KCl) and a proteome adapted to these high salt conditions.

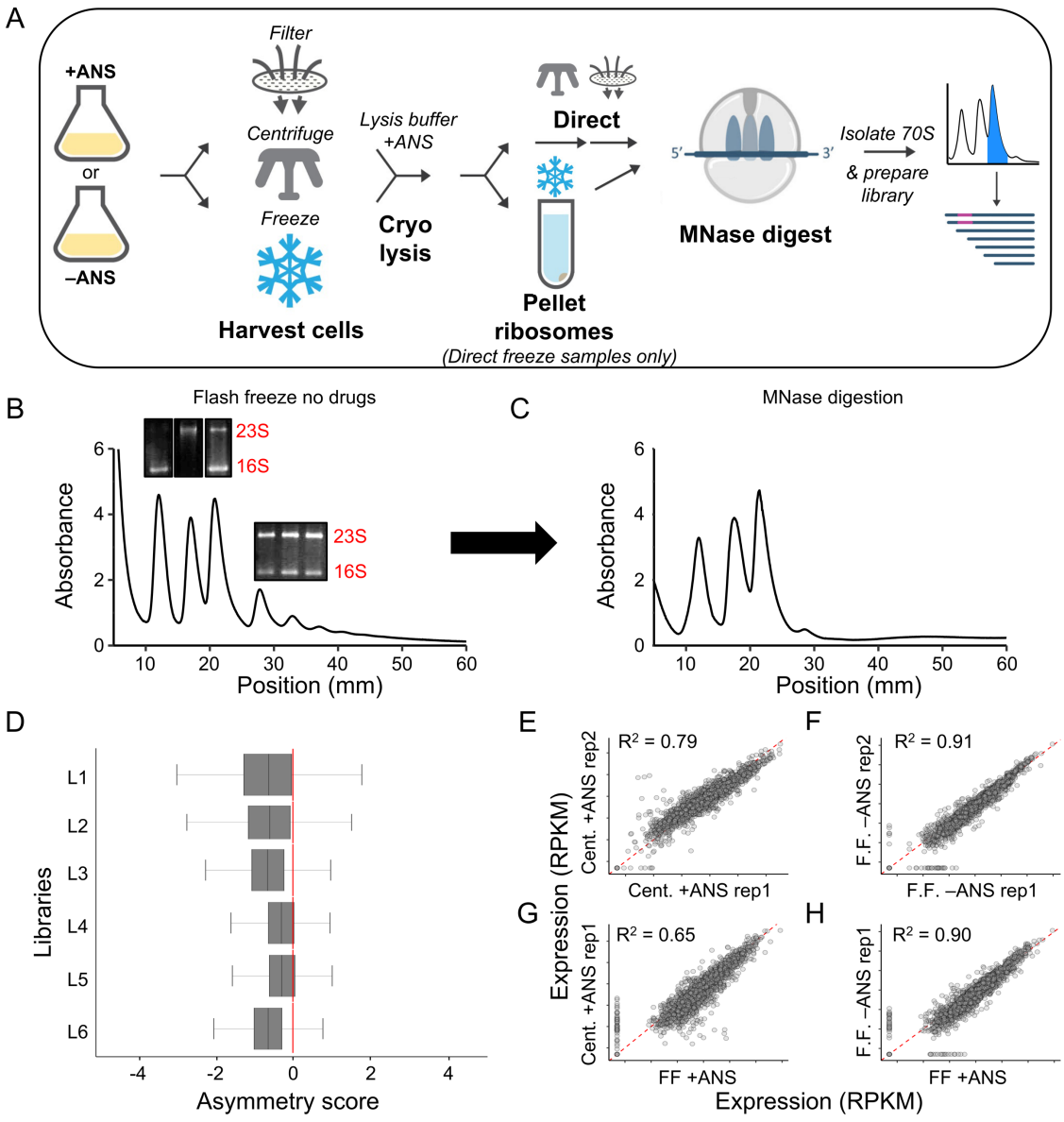
A
+ANS
or
–ANS
Filter
Centrifuge
Freeze
Harvest cells
Lysis buffer
+ANS
Cryo
lysis
Direct
Pellet
ribosomes
(Direct freeze samples only)
5′
3′
MNase digest
Isolate 70S
& prepare
library
B
Flash freeze no drugs
23S
16S
23S
16S
Absorbance
Position (mm)
C
MNase digestion
Absorbance
Position (mm)
D
Libraries
L1
L2
L3
L4
L5
L6
Asymmetry score
E
R² = 0.79
Cent. +ANS rep2
Cent. +ANS rep1
F
R² = 0.91
F.F. –ANS rep2
F.F. –ANS rep1
G
R² = 0.65
Cent. +ANS rep1
FF +ANS
H
R² = 0.90
F.F. –ANS rep1
FF +ANS
Expression (RPKM)
Expression (RPKM)

Figure 3-1: Evaluation of ribosome profiling protocols for *H. volcanii*. (A) Schematic of the ribosome profiling methods used in this study. Sucrose gradients of ribosomes from cells harvested by direct flash freezing in liquid nitrogen with no drugs added to cultures showing (A) intact monosome and polysome fractions, and (B) successful collapse of polysomes into monosomes after MNase digestion. Inset: Total RNA agarose gels showing rRNA from various peaks corresponding to 30S, 50S, 70S, and polysome fractions. (C) Asymmetry scores were calculated as the normalized ribosome density upstream of the halfway point of an ORF over the ribosome density downstream of the halfway point of an ORF. L1= Cent + ANS rep1, L2= Cent + ANS rep2, L3= FF + ANS, L4= FF no drug rep. 1, L5= FF no drug rep. 2, L6= FF + Harringtonine). (D-G) Correlation plots of translation expression (in rpkm) for ribosome profiling libraries of various conditions. ANS, anisomycin; Cent, centrifugation; FF, flash-freezing; rep, biological replicate.

In ribosome profiling studies in bacteria and yeast, cultures are commonly harvested by either filtration or centrifugation. After filtration (for < 1 min), cells were collected from the filter, flash-frozen in liquid nitrogen, and lysed in a cryo-mill. In *H. volcanii*, the filtration procedure did not yield intact ribosomes (**Fig. 3-S1A**). Using centrifugation at room temperature for 3 min, we observed 30S/50S subunits but very few translating ribosomes (70S monosomes and polysomes, **Fig. 3-S1B**). This result suggests that *H. volcanii* ribosomes ran off messages during centrifugation, perhaps due to the drop in temperature from 42 °C to 25 °C (Zhang et al., 2018). To arrest ribosomes prior to centrifugation, we tested two antibiotics known to inhibit translational elongation,

anisomycin and cycloheximide. We found that adding anisomycin (ANS) to the culture prior to centrifugation halted growth and stabilized translating ribosomes during the centrifugation step (**Fig. 3-S1C & Fig. 3-S2**). Centrifugation in the presence of ANS, therefore, offered one potential solution for harvesting *H. volcanii* cultures for ribosome profiling.

We also tested a third method for harvesting cells that we recently developed for *E. coli* (Mohammad et al., 2019). In this method, cultures were sprayed directly into liquid nitrogen. Flash freezing the cells in the media arrests translation immediately without the use of antibiotics. The frozen cultures were then lysed in a cryo-mill and, in an extra step, ribosomes were pelleted and resuspended into a high-salt lysis buffer (3.4 M KCl, 500 mM $MgCl_2$, 10 mM Tris-HCl pH 7.5) with 100 μg/ml ANS to inhibit translation in the lysate. We found that this flash-freezing method yielded the highest level of translating ribosomes (70S monosomes and polysomes) relative to free 30S and 50S subunits (**compare Fig. 3-1B with Fig. 3-S1C**). These data show that the flash-freezing protocol developed for bacteria yielded robust polysomes from *H. volcanii* cultures and indicated that ribosomes were arrested and remained intact throughout the purification procedure.

Due to high KCl concentration in the lysis buffer, RNA digestion with micrococcal nuclease (MNase) was inefficient and lowering KCl concentrations to 1 M or 2 M destabilized 70S ribosomes and favored subunit splitting (**Fig. 3-S3**). Reasoning that high concentrations of KCl might prevent the catalytic metal ion Ca^{2+} from binding in the active site of MNase, we titrated $CaCl_2$ in the RNA digestion reaction. We found that with 50 mM $CaCl_2$ in the lysis buffer, MNase was able to collapse polysomes into

monosomes (**Fig. 3-1C**). After purifying 70S monosomes from a sucrose gradient, we isolated 10 – 45 nt mRNA fragments and prepared cDNA libraries for deep sequencing.

Prior studies in yeast and *E. coli* have shown that adding antibiotics to the media introduces artifacts in ribosome profiling studies by trapping ribosomes near the 5'-end of ORFs as elongation is blocked while initiation continues (Gerashchenko & Gladyshev, 2014; Mohammad et al., 2019). To test whether the same problem occurs in *H. volcanii*, we sequenced libraries prepared by centrifugation with ANS pre-treatment (**Fig. 3-1D, L1 and L2**) and by flash-freezing cultures with (**L3**) or without (**L4 and L5**) ANS pre-treatment. We then computed asymmetry scores for each gene, taking the ratio of ribosome density in the second half of the gene over the density in the first half ($\log_2$ transformed which can yield negative values). If the asymmetry score is less than zero, ribosome density is accumulating at the 5'-end of the ORF. We found that cultures pre-treated with ANS prior to harvesting (**Fig. 3-1D, L1-3**) had significantly lower asymmetry scores than those not treated with ANS (**L4 and L5**), regardless of the cell harvesting method, centrifugation or flash-freezing. These results indicate that adding antibiotics to the culture drastically altered the ribosome position along ORFs.

We further investigated the effects of ANS treatment by determining the relative level of gene translation (reported in reads per kilobase per million mapped reads, or RPKM) in libraries of biological replicates prepared by centrifugation and by flash-freezing. Translational levels for each gene were more highly correlated between the two replicates harvested by flash freezing (**R^2 = 0.91, Fig. 3-1F**) than between those harvested by centrifugation (**R^2 = 0.79, Fig. 3-1E**), indicating that flash freezing is a more reproducible harvesting method. This correlation was reduced when comparing

centrifugation and flash-freezing methods in cultures pre-treated with ANS (**$R^2 < 0.65$, Fig. 3-1G**). In contrast, there was a strong correlation between flash freezing samples, even when one culture was treated with ANS and the other was not (**R^2=0.90, Fig. 3-1H**), indicating that the harvesting method (i.e. centrifugation) is the critical difference that alters translational levels and that ANS treatment is less of a factor. Ultimately, we concluded that flash freezing without ANS pre-treatment introduced the fewest artifacts and thus all subsequent libraries described below were prepared with this method.

27 nt RPFs reveal the reading frame of the ribosome in H. volcanii

Isolating RPFs between 10 – 45 nt, we consistently found in our ribosome profiling libraries that the largest number of RPFs were 27 nt (**Fig. 3-2A**). The distribution of RPFs lengths was reproducible, with only subtle differences between libraries regardless of how they were prepared (by flash freezing or centrifugation, with or without ANS or HHT pre-treatment). The predominant 27 nt footprints in *H. volcanii* (**Fig. 3--2B, green**) were only slightly shorter than the predominant footprints in eukaryotes (28 – 30 nt, red, orange, and brown). In contrast, most RPFs in *E. coli* are shorter, around 24 nt (blue). In addition, we observed that the distributions for all these species have more than one peak. In yeast and mammalian cells, there are two distributions of RPF lengths, one centered at 28 nt and the other at 21 nt. These two peaks represent two different conformational states of elongating ribosomes (Lareau et al., 2014; Wu et al., 2019). In all of our *H. volcanii* ribosome profiling libraries, we also observe a peak of short RPFs (**Fig. 3-2B**), but as described below, they differ from the short reads in eukaryotes.

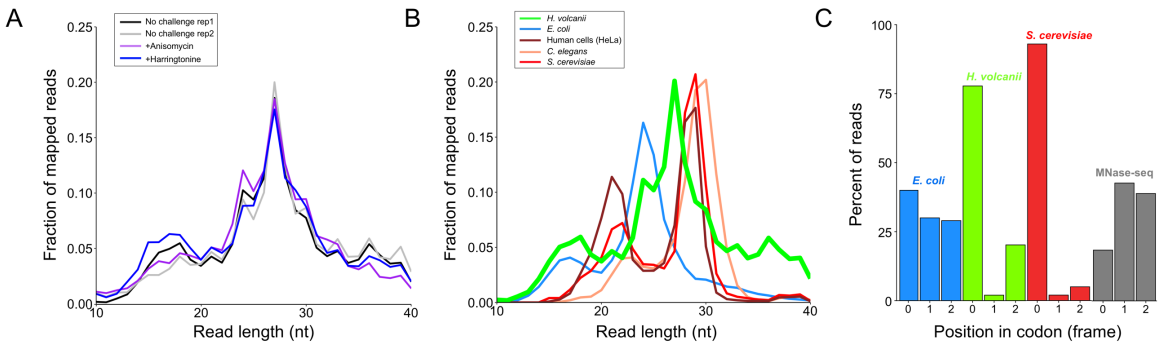

Figure 3-2: Length distribution of ribosome footprints in *H. volcanii.* (A) Fraction of reads of various lengths in *H. volcanii* ribosome profiling libraries from cultures harvested by flash-freezing without (2 biological replicates) and with anisomycin (ANS) treatment prior to harvesting. (B) Fraction of reads of various lengths in ribosome profiling libraries for *E. coli, S. cerevisiae*, *C. elegans*, mammalian cells (HeLa), and *H. volcanii* cells. (C) Mapped distribution of the elongating footprints to the position in codon (0, +1, or +2) for *E. coli*, *H. volcanii*, *S. cerevisiae*, and total *H. volcanii* RNA treated with MNase (MNase-seq).

One of the hallmarks of active translation that is captured by ribosome profiling is that elongation occurs one codon at a time. Because 28 nt RPFs in yeast are fully trimmed by nucleases back to the 5'- and 3'-boundaries of the ribosome, they reveal the position of the ribosome on an mRNA at higher resolution than RPFs of other lengths. 90% of 28 nt RPFs in yeast map to the first nt in codons, evidence of strong three nt periodicity (**Fig. 3-2C, red**). Likewise, we observe in *H. volcanii* that the majority of 27 nt RFPs (> 75%, green) map to the first nt of codons (**Fig. 3-2C, green**), demonstrating that these RFPs arise from elongating ribosomes moving one codon at a time along mRNAs. These results contrast sharply with those in bacteria: the periodicity in *E. coli* is much weaker with a nearly equal distribution of mapped footprints to the first (40%), second (30%), and third nt (30%) of codons (**Fig. 3-2C, blue**). In *Hv*, 27 nt RPFs alone

show the strongest periodicity compared to all other RPF lengths (**Fig. 3-S4A**) and compared to the periodicity of all RPFs averaged together (**Fig. 3-S4B**). It is known that MNase can yield tight distributions of footprints and strong periodicity in yeast ribosome profiling studies (Gerashchenko & Gladyshev, 2017). To confirm that the three nt periodicity in *H. volcanii* was due to ribosome's reading frame, we performed RNA-seq using MNase to fragment total RNA, followed by the same steps for library preparation as in ribosome profiling. In the absence of ribosomes, this RNA-seq protocol with MNase yielded a relatively equal distribution of footprints at all three positions of codons (**Fig. 3-2C, gray**). The sharp contrast between the lack of periodicity in this RNA-seq protocol and the strong periodicity seen in 27 nt RPFs from ribosome profiling validates that we are accurately following translation at high resolution in *H. volcanii*.

Hallmarks of actively initiating and elongating ribosomes identified in H. volcanii

One of the many strengths of ribosome profiling lies in the averaging of ribosome density across thousands of genes, aligned at their start codons, to perform "meta-gene" analyses. The position of the ribosome on the mRNA can be assigned using either the 5'-end of the footprint, as is commonly done in eukaryotes, or the 3'-end of the footprint, giving better resolution in bacteria (Mohammad et al., 2019; Nakahigashi et al., 2014). Representative examples of meta-gene plots for *H. volcanii* are shown in Fig. 3, using ribosome density with positions assigned either at the 5'-end (**Fig. 3-3A**) or 3'-end of footprints (**Fig. 3-3B**). These meta-gene plots revealed high ribosome density and three-nucleotide periodicity within ORFs, and lower ribosome density and no periodicity in untranslated regions (**Fig. 3-3A, B**). Heatmaps of average ribosome density, separated by footprint size, for both 5'- and 3'-mapped data demonstrated a tight density of 27 nt

footprints along ORFs with strong periodicity relative to all other footprint lengths (**Fig. 3-3A, B**), consistent with the reading frame analysis shown in **Fig. 3-2C**. These findings are consistent with hallmarks of elongating ribosomes observed in previous ribosome profiling studies.

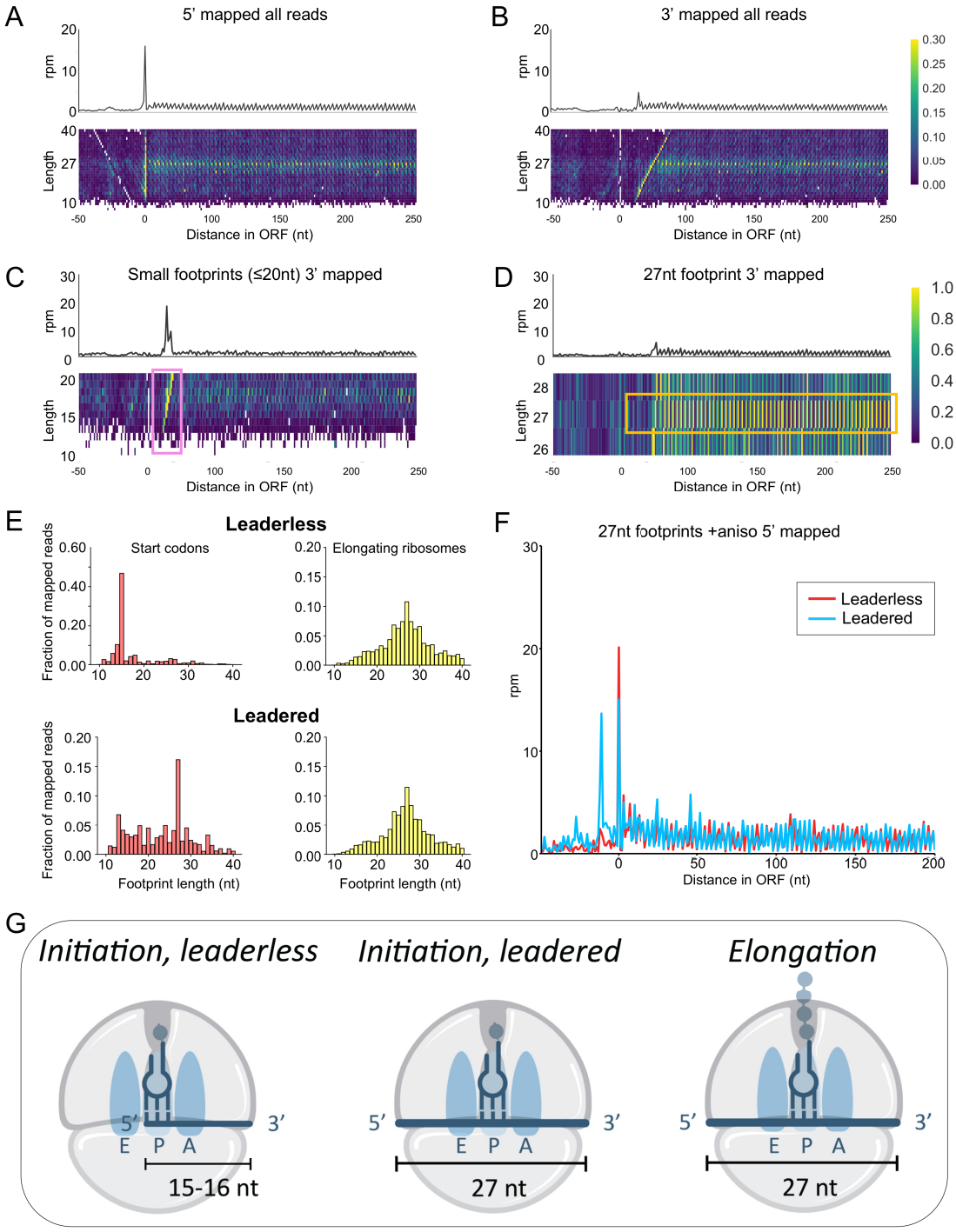

Figure 3-3: Different length ribosome footprints represent different translational states. Meta-gene plots, from flash frozen and no drug exposure *H. volcanii* cultures, of ribosome density mapped using the (A) 5'-end and (B) 3'-end of ribosome footprints (rpm = reads per million). Heatmaps of footprint density separated by length are aligned underneath the meta-gene plots. (C) 3'-mapped meta-gene analysis of small footprints (10-20 nt) and (D) of 27 nt footprints from flash frozen and no drug cultures. (E) Footprint length distribution of 5'-mapped ribosome density for initiating ribosomes at the start codon (at P-site) in leadered (5' UTR ≥ 10 nt) and leaderless (0 nt) transcripts (left, red) and elongating ribosomes within the open reading frames (right, yellow). (F) 27 nt footprints on leadered (5' UTR ≥ 10 nt) and leaderless (0 nt) messenger RNAs from flash frozen and anisomycin (ANS) treatment prior to harvesting. (G) A model for footprint lengths corresponding to ribosome translation states. Small footprints

Another common feature of metagene plots is a peak of ribosome density at the start codon that reflects a delay of newly initiated 70S ribosomes in entering the elongation cycle. Although we observed a strong peak in 5'-mapped data (**Fig. 3-3A**), the 5'-ends of RPFs were enriched at zero, the first nt in the AUG start codon, rather than enriched at 12 nt upstream of the start codon (the distance between the 5'-boundary of the ribosome and a start codon bound in the P site). In a heatmap of average ribosome density broken down by footprint size, there was a vertical line of high density at this same position, indicating that the 5'-end of footprints line up with the A in the AUG start codon regardless of the footprint length. In contrast, in the 3'-mapped data (**Fig. 3-3B**), the peak near the start codon is significantly lower and shifted downstream. In the

associated heatmap broken down by footprint length, there was a diagonal line of high density because the 3'-end of footprints moves downstream as a function of footprint length. Regardless of how we plotted these data, we consistently found an enrichment of footprints whose 5'-ends lies at the first nt of the start codon. We investigated the difference between these footprint sizes further by plotting separately the distribution of small footprints (10-20 nt) and long footprints (26-28 nt) along ORFs using 3'-mapped meta-gene analysis (**Fig. 3-3C-D**). The distribution of small footprints showed a bias towards the translation start site, with a peak density at 16 nt, and limited periodicity along ORFs for *H. volcanii* cells with no drug treatment (**Fig. 3-3C**). In the longer footprint distribution with no drug treatment, the TSS bias was absent and periodicity along the ORF continued downstream, with a peak density at 27 nt (**Fig. 3-3D**).

More than 70% of *H. volcanii* mRNAs have been reported to be leaderless (Babski et al., 2016), starting with the AUG start codon. We reannotated the *H. volcanii* transcriptome with previously published RNA-seq data (Gelsinger & DiRuggiero, 2018), confirming that the majority of transcripts in *H. volcanii* were indeed leaderless (**Fig. 3-S5A**). As expected, we observed far fewer reads in RNA-seq data in the 5'-UTR of leaderless transcripts than in leadered transcripts (**Fig. 3-S5B**). As a result, the enrichment of RPFs at the AUG in **Fig. 3-3A** likely arises from the fact that these fragments already have one defined end (the 5'-end), resulting in a cloning bias. In contrast, fragments in the middle of an mRNA have to be cleaved on both ends to be isolated within the 10 – 45 nt cutoff and included in the library (Lalanne et al., 2018). We conclude that the strong peak in **Fig. 3-3A** and **3-3C** at least partially reflects an artifact of conducting ribosome profiling on a transcriptome high in leaderless transcripts.

Next, we asked how ribosome footprints differ at start sites on leadered vs leaderless transcripts in *H. volcanii*. As shown in **Fig. 3-3G**, we hypothesized that newly-initiated ribosomes on leadered transcripts would protect longer footprints (27 nt) than newly-initiated ribosomes on leaderless transcripts. If the 5'-end of the transcript (the A in AUG) lies in the ribosomal P site, the ribosome should only protect 15-16 nt from nuclease digestion (the distance from the P-site codon to the 3'-boundary of the ribosome). To test this hypothesis, we used the data from L3 (cells treated with ANS) where we saw the most ribosome density at start codons (**Fig. 3-S6**). For RPFs with start codons in the P site on leaderless transcripts, we observed a strong enrichment of 16 nt RPFs and very few 27 nt RPFs, as expected (**Fig. 3-3E, Leaderless**). In contrast, in leadered transcripts, there were significantly more 27 nt RPFs at start codons (**Fig. 3-3E, Leadered**) because the ribosome protects additional mRNA upstream of the start codon (**as shown in Fig. 3-3G**). In both leadered and leaderless transcripts, RPFs mapping within the ORF tended to be long, with a peak at 27 nt, corresponding to elongating ribosomes (**Fig. 3-3E**).

Another way to visualize these differences at start codons is to plot average ribosome density for only leadered or leaderless genes (**Fig. 3-3F**). Using exclusively 27 nt RPFs, we found that their 5'-ends lie at the TSS (zero) in leaderless mRNAs, consistent with two possibilities: (i) the artefactual cloning bias discussed above, and (ii) footprints from newly-initiated ribosomes with no mRNA upstream of the start codon. We cannot distinguish between these two possibilities on leaderless transcripts, but it is likely that both contributed to the signal, because we also observed these on leadered transcripts, where we can differentiate between them. In leadered transcripts, the cloning

bias also yields a peak at the TSS (just like leaderless transcripts). But importantly, there is also a strong peak 12 nt upstream of the TSS, consistent with the 5'-end of RPFs with the start codon positioned in the P site (**Fig. 3-3F**). This peak corresponds to newly initiated ribosomes.

In summary, we found that elongating ribosomes in *H. volcanii* protect a 27 nt footprint on both leadered and leaderless mRNAs (**Fig. 3-3G**). More importantly, our analysis indicated that short footprints (< 20 nt) were strongly enriched at the TSS of leaderless transcripts, either by 5'-cloning artifacts or by bona fide initiation, whereas the 27 nt footprint peak upstream of start codons on leadered mRNAs could be attributed to initiating ribosomes. These observations provide a way to use newly-initiated ribosomes to identify novel initiation sites and better annotate the *H. volcanii* genome, as described below.

Ribosome profiling detects codon specific translational pauses in H. volcanii

Ribosome profiling studies are able to capture ribosome pausing because when a specific codon is translated more slowly on average, the increased ribosome occupancy indicates that there are more RPFs associated with that codon genome-wide. To test if ribosome pausing could be detected at high resolution in *H. volcanii*, we treated cells with a compound designed to induce pauses at specific codons. Serine hydroxamate (SHX) is an amino acid analog that acts as a competitive inhibitor of seryl-tRNA synthetase in *E. coli* (Mohammad et al., 2019; Tosa & Pizer, 1971). We found that 6 mM SHX halted the growth of *H. volcanii*, indicating that the inhibitor could disrupt its metabolism (**Fig. 3-S2**). Using ribosome profiling data from cells treated with SHX, we computed pause scores for all 61 sense codons. To calculate pause scores we normalized

the read count at each nt of a gene by dividing by the mean read count for the gene. For each codon, we calculated the mean value including reads from all three nt. Average pause scores were calculated using these values from all instances of the codon or amino acid of interest. Pause scores calculated for the A site used a −11 nt shift; P- and E-site pause scores used a shift of −14 and −17, respectively. When codons were positioned in the E site or the P site of the ribosome, we observed only small deviations from the expected value of 1, indicating little or no pausing (**Fig. 3-4A**). For the ribosomal A site, however, where decoding takes place, we observed very strong pauses (score = 7) on the codon AUG, which encodes methionine (these calculations did not include start codons). There was also a weaker pause at cysteine codons in the A site (score = 2). These observations are consistent with pausing during decoding as the ribosome waits for a cognate Met or Cys aminoacyl-tRNA that is in low concentration. Surprisingly, there was no evidence of the expected pauses at Ser codons. Limiting our analysis to the 27 nt RPFs revealed the contrast between the Met and Ser codons even more clearly in plots of average ribosome density centered on the codon of interest (**Fig. 3-4B**).). Importantly, when conducting the same pause score analysis on no drug treated ribosome profiling libraries, we did not find pausing on Met nor Ser codons (**Fig. 3-4A**). This additional analysis confirmed that the methionine pauses were specific to the A site of ribosomes due to SHX exposure and confirmed the lack of pauses at or around Ser codons.

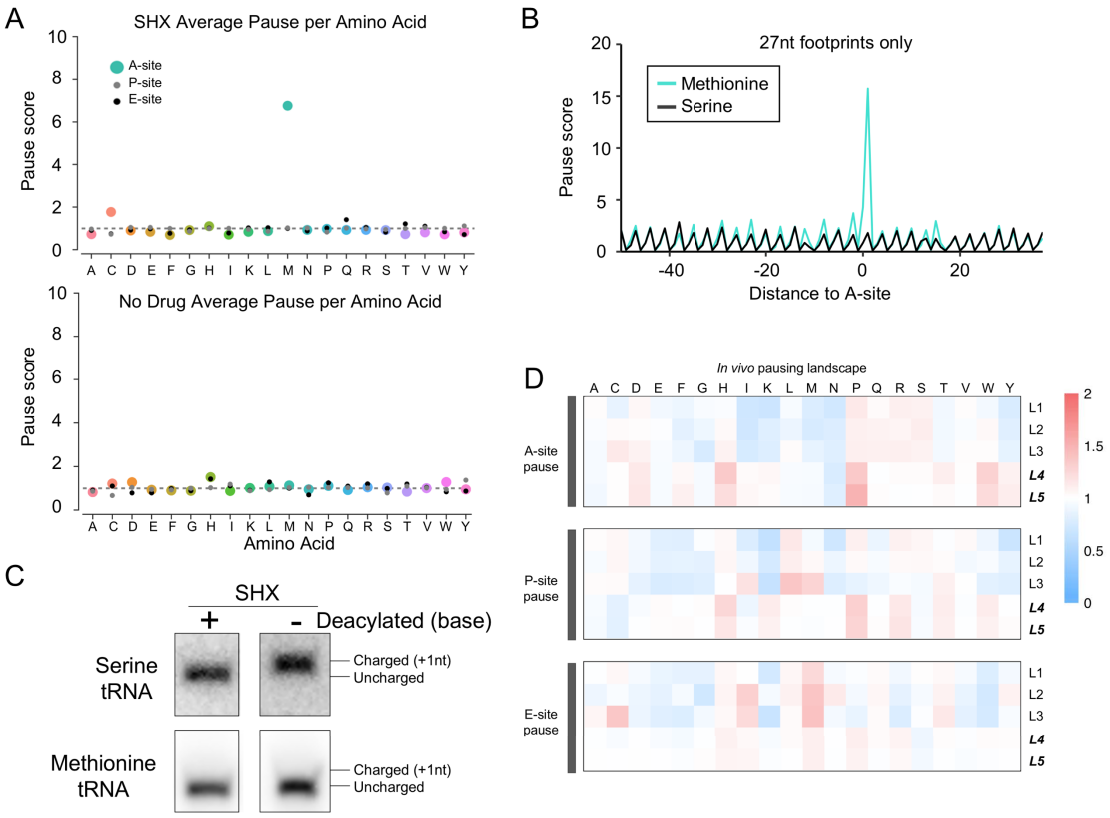

Figure 3-4: *H. volcanii* global ribosome pausing. (A) Pause scores for all footprints were calculated for each codon and averaged per amino acid at the A- (colored large dot), P- (small gray dot), and E- (small black dot) sites of ribosomes for serine hydroxamate (SHX) treated cells. Ribosome sites were calculated from the 3' of elongating footprints. (B) Meta-gene plots of ribosome density for elongating footprints (27 nt) for methionine and serine codons. (C) Northern blots of charged and uncharged Met or Ser tRNAs from total RNA of SHX-treated cells. Deacylation treatment is indicated by + and charged/uncharged tRNAs are marked. (D) Comparative heatmap of pause scores for all amino acids under several conditions for A-site, P-site, and E-site of the ribosome.

The unexpected observation of strong Met pauses in *H. volcanii* led us to ask if SHX lowers the concentration of Met-tRNA rather than Ser-tRNA as it does in *E. coli*. We extracted total tRNA from *H. volcanii* treated with SHX and, as a control, pre-treated

one aliquot with mild base to deacylate all the tRNAs. We then used periodate oxidation and β-elimination to distinguish between charged and uncharged tRNA in these samples. Because uncharged tRNAs are selectively oxidized by periodate, they are one nt shorter than charged tRNAs, allowing their resolution by PAGE and northern blotting using tRNA-specific probes. We observed that in SHX-treated cells, Ser-tRNA (AGC) was fully charged and therefore longer than the deacylated control (**Fig. 3-4C**). In contrast, Met-tRNA was uncharged and, therefore, ran at the same size as the deacylated control (**Fig. 3-4C**). These results indicate that SHX reduced the level of charged Met-tRNA in *H. volcanii* and not Ser-tRNA, consistent with the A-site pauses observed in the ribosome profiling data. In addition, we found that genes involved in KEGG pathways related to the biosynthesis of methionine and cysteine were up-regulated in ribosome profiling data, reflecting higher levels of expression during SHX treatment; in contrast, serine biosynthesis was unaffected. At this point it remains unclear whether charging of $tRNA^{Met}$ is inhibited directly by SHX or if it acts by blocking the biosynthesis of Met which typically involves Ser as a methyl donor.

Our observations of pauses induced by SHX gave us confidence that we could assess the *in vivo* pausing landscape in *H. volcanii* cells with and without ANS pre-treatment during harvesting. Heatmaps reflecting pause scores for codons in the A-, P-, and E-sites of elongating ribosomes are shown in **Fig. 3-4D**. In untreated cells (**L4 and L5**), the strongest pauses at the A and P sites occurred at Pro codons, consistent with the known biochemistry of Pro as both a poor peptidyl donor and acceptor in the active site of the ribosome. Comparison of the untreated and ANS treated samples confirmed what was previously shown in yeast and in *E. coli* (Hussmann et al., 2015; Mohammad et al.,

2019; Sothiselvam et al., 2014), that pre-treatment of cultures with elongation inhibitors (**L1-L3**) obscures the natural pausing landscape observed in **L4 and L5**. These data once again highlight the importance of harvesting cells without pre-treatment with antibiotics.

High-throughput identification of translation start sites

The translation inhibitors harringtonine (HHT) and lactimidomycin have been used in eukaryotic cells to enrich ribosomes at initiation sites by inhibiting elongation during the first rounds of peptide-bond formation after subunit joining (FRESNO et al., 1977; Robert et al., 2009). Importantly, these compounds are less active against elongating ribosomes are not inhibited by these antibiotics, allowing them to continue elongation and terminate normally at a stop codon (FRESNO et al., 1977; Ingolia et al., 2011; Lee et al., 2012; Robert et al., 2009). Although these compounds have proved useful in eukaryotic studies, an important caveat is that their treatment can alter the relative level of initiation at start codons. We reasoned that we could identify translational start sites globally in *H. volcanii* by leveraging the effects of HHT. We determined that 1 mg/mL HHT prevented growth in *H. volcanii* (**Fig. 3-S2**), a concentration that is an order of magnitude higher than required for eukaryotes. We then conducted ribosome profiling with cells treated with HHT. In metagene plots aligned at start codons, we found that ribosomes strongly accumulated at annotated TSS in the presence of HHT compared with untreated samples (**Fig. 3-5A**). Another line of evidence supporting the effectiveness of HHT treatment was that the asymmetry score in HHT-treated cells was biased towards the first half of ORFs (**library L6, Fig. 3-1C**).

We then leveraged HHT-treatment to comprehensively identify initiation sites in *H. volcanii* using previously established computational methods (Meydan et al., 2019;

Weaver et al., 2019). In brief, we identified all potential ORFs, asked if peaks of ribosome density aligned near the start codon of each potential ORF (± 3 nt), verified that these ORFs were translated in profiling data without HHT treatment (≥ 1 rpkm ribosome density), and assigned the type of TSS based on the gene annotation (NCBI RefSeq GCF_000025685.1_ASM2568v1). Among the ORFs passing our analyses, 1,413 were detected out of the 2,780 previously annotated protein coding genes on the main chromosome. In several examples, the ribosome density on these genes was sharply increased at the start site in comparison to the rest of the ORF in HHT-treated but not in untreated samples (**Fig. 3-5B**). Yet there were other genes with little or no enrichment at the start codons. We cannot explain why some genes appear to be more sensitive to HHT than others. To increase the chances of finding bona fide start sites, we calculated the enrichment of ribosome density at the TSS for each potential ORF compared with the downstream density within the ORF. Candidates with high relative density (≥ 0.1) were considered HHT sensitive; 188 of the 1,413 previously detected annotated start sites were above this stringent threshold.

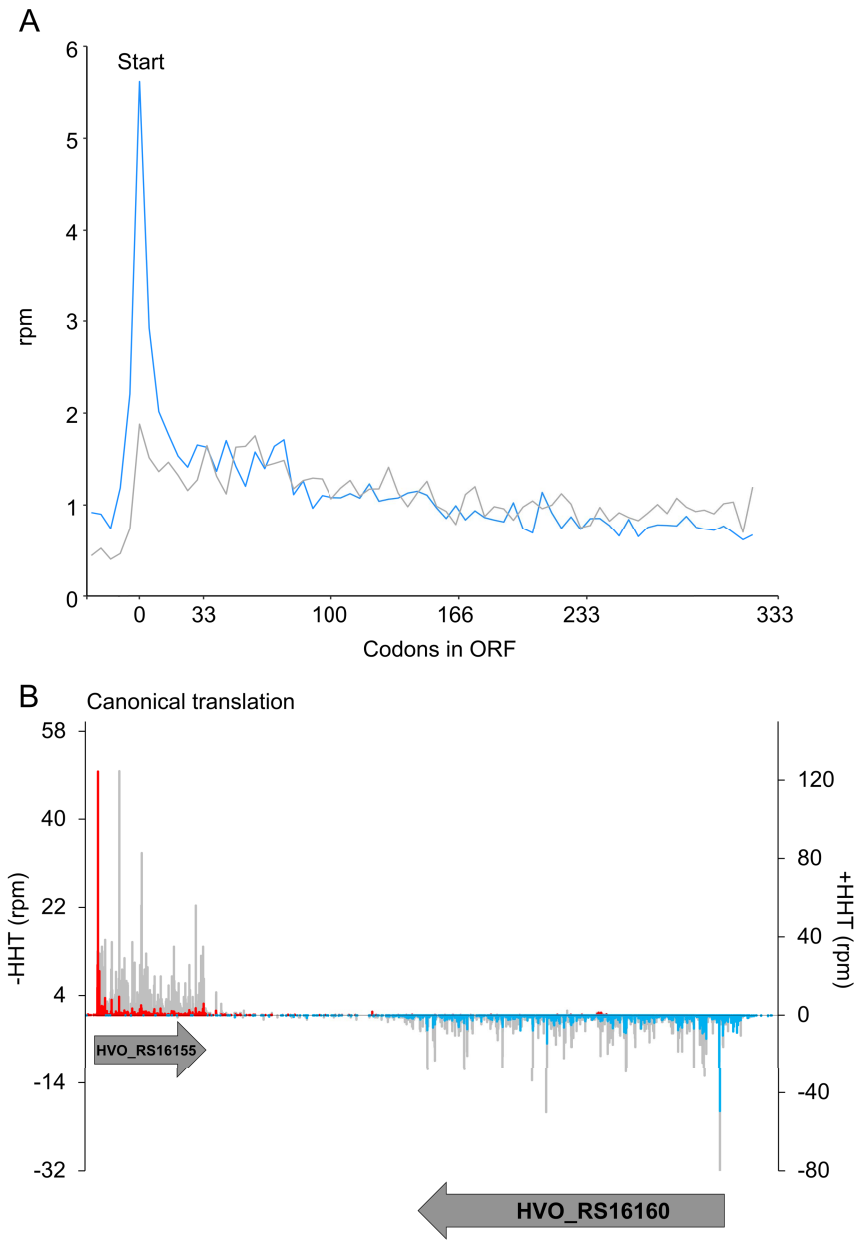

Figure 3-5: Harringtonine (HHT) locks ribosomes at translation start sites. (A) Meta-gene analysis plot of ribosome density along ORFs 1 kb downstream from the TSS. The blue line is HHT treatment and the gray line is no-drug treatment. (B) Representative plots of translation start site identification in annotated genes. Blue bars are HHT ribosome density on the minus strand, red bars are on the plus strand. Gray bars are ribosome density from untreated samples on both strands.

A hallmark of translation is the binding of ribosomes onto coding sequences (CDS). We found that the majority (79%) of footprints during no drug conditions map to known CDS, while a minority map to 3' and 5' UTRs (**Fig. 3-S7A**). A minority of footprints mapped to intergenic regions, which may correspond to unannotated novel proteins (Fig S6A). Unexpectedly, 14% of footprints map to non-coding RNAs (ncRNAs) such as RNaseP, signal recognition particle (SRP), and small non-coding RNAs (sRNAs) previously identified (**Fig. 3-S7A**). A characteristic of these ncRNAs is the absence of ORF potential. In order to validate the 14% of footprints that map to ncRNAs we tested whether ribosome footprint density on ncRNAs reflected *in vivo* translation. We surmised that if ribosome density on ncRNAs in ribosome profiling libraries altered with HHT-treatment compared to no drug-treatment that this indicated active translation on the ncRNAs. Carrying out this analysis on a *bona fide* ncRNA, RNaseP, we observed no differences in ribosome density along the transcript in both HHT and no drug conditions (**Fig. 3-S7B**). This held true for 95% of all ncRNAs with ribosome density. In contrast, a small proportion of intergenic sRNAs (6) demonstrated altered ribosome density with HHT treatment indicating active translation, although the potential translated ORFs were small (< 30 amino acids) and occupied only a portion of the length of ncRNA transcript (**Fig. 3-S7C**).

Characterization of putative novel and putative alternative initiation

Using the same stringent parameters as for the 188 HHT-sensitive TSS, we analyzed the other TSS in our dataset that have not previously been annotated as start sites. We detected 160 novel TSS and classified them as putative alternative TSS (paTSS) based on their relationship to the annotated ORF (**Fig. 3-6A**). We identified both putative

small ORFs less than 50 codons (smORFs, 68 of 160) and putative novel unannotated TSS that encode large ORFs (18 of 160) in both intergenic regions and antisense to annotated genes. In addition, paTSS were identified upstream of annotated TSS (N-terminal extension, 31 of 160), or internal to annotated ORFs either in-frame (27 of 160) or out-of-frame (16 of 160). We observed that many paTSS started with AUG, GUG, and UUG, while ~30% of paTSS started at non-canonical start sites that were near-cognate codons to AUG (**Figure 3-S8)**. In order to limit false positives, we selected only paTSS that started with AUG or the other known alternate start codons GUG and UUG for further analysis (**Fig. 3-6B**).

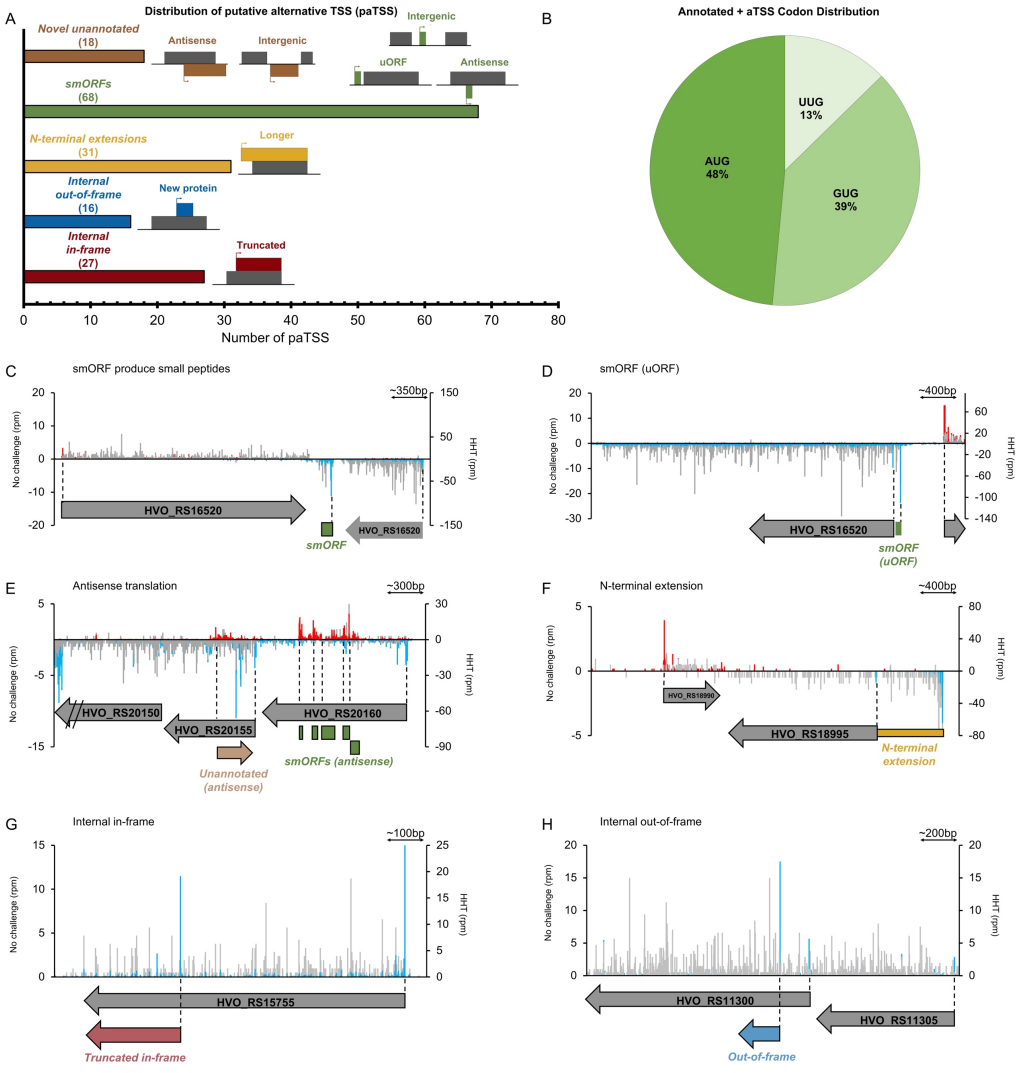

Figure 3-6: Identification of alternative translation initiation sites (aTSS). (A) Distribution of HHT-identified aTSS. (B) Distribution of initiation codons for HHT-identified aTSS. (C-H) Representative examples of aTSS identified in HHT-treated cells. (C) Intergenic smORF, (D) upstream smORF (uORF), (E) antisense translation initiation, (F) N-terminal extensions, (G) internal in-frame, and (H) internal out-of-frame initiation. Plus strand is plotted with positive values and minus strand with negative values. Untreated samples are shown with gray bars on both strands, HHT conditions on the plus strand are red bars, and blue bars are HHT conditions on the minus strand.

Small proteins have recently been found to be important players in a variety of processes in both Eukaryotes and Bacteria (Andrews & Rothnagel, 2014; Galindo et al., 2007; Hanada et al., 2013; Hemm et al., 2008; Hobbs et al., 2012; T Kondo et al., 2010; Kondo et al., 2007; Storz et al., 2014; Weaver et al., 2019) but have not been well characterized in Archaea. Using our ribosome profiling approach, we found many paTSS that were predicted to produce small proteins. smORFs were found in intergenic regions (**Fig. 3-6C**), as distinct proteins in 5'-UTRs of mRNAs (upstream ORFs) (**Fig. 3-6D**), and antisense to coding regions in clusters (**Fig. 3-6E**). Many of these putative smORFs were highly conserved in other haloarchaea, suggesting their importance and functional potential.

We also observed N-terminal extensions of annotated genes, with enrichment of footprints at an upstream TSS in HHT-treated samples followed by elongating footprints extending to annotated ORF. In many cases peaks were also observed at the annotated TSS, suggesting that two distinct proteins may be produced from a single gene (**Fig. 3-6A**). For example, translation of a DEAD/DEAH box helicase likely initiated at its

annotated TSS as well as and potentially as an alternate TSS 273 nt upstream, allowing for the possibility of encoding two isoforms of this helicase (**Fig. 3-6F**).

Internal translational initiation could potentially produce N-terminal truncations (in-frame) or completely new proteins (out-of-frame) within a longer ORF, increasing the repertoire of proteins in a compact genome. One example is a gene encoding deoxyhypusine synthase that likely initiated at its annotated start site (AUG) and from an internal TSS 738 nt downstream in the same reading frame (**Fig. 3-6G**). The annotated deoxyhypusine synthase is a 359 amino acid long protein involved in post-translational modifications. The putative internal TSS potentially produces an N-terminally truncated protein only 113 amino acids long, losing the NAD/FAD-binding domain while retaining the transmembrane helix of deoxyhypusine synthase. In contrast, glutamate-5-semialdehyde dehydrogenase, involved in amino acid biosynthesis, initiated at the annotated TSS and also from an internal TSS 203 nt downstream in a different reading frame (**Fig. 3-6H**). This putative alternative frame could potentially produce a 75 amino acid protein as opposed to the annotated 444 amino acid protein. This novel internal out-of-frame protein was predicted to be highly disordered (89% of sequence) but had a single significant domain hit using blastp against the NCBI non-redundant (nr) database (version 2019/03/22) which was an acyl transferase domain (DELTA-BLAST, e-val = 2e-04, SEB92775.1) involved in polyketide synthesis.

Genes involved in the biosynthesis of amino acids and secondary metabolites (p=6.4E-4) and GTPase activity (p=3.7E-2) were enriched in the set of genes with N-terminal extensions, the latter including the cell division protein FtsZ and elongation factor 1-alpha. The internal TSS demonstrated a GO enrichment in universal stress

proteins (UspA) and cell redox homeostasis (p=4.5E-2) for putative N-terminally truncated genes but no significant GO enrichment for out-of-frame TSS.

Exploring translation of putative alternative TSS

Harringtonine treatment allowed us to use ribosome profiling to predict TSS at the global level, but the data could provide insight into the relative translational efficiencies of annotated and putative alternative TSS. We coupled standard ribosome profiling with RNA-seq to determine the translation efficiency for ORFs corresponding to the experimentally identified TSS in this study. Translation efficiency is a ratiometric estimation of the number of ribosomes per mRNA, where increases and decreases of efficiency are deviations from the baseline of 1:1 RPF and transcript levels. It is important to note that for paTSS we calculated translation efficiency for only the predicted ORF, or N-terminal extension, and removed the other regions of the annotated ORF. While the majority of paTSS and their corresponding genes were translated at levels comparable to their transcript levels, we observed outliers with high and low translation efficiency through a pair-wise comparison (**Fig. 3-7A**). Because the 3' UTRs of genes are not translated, their low translation efficiency values (**Fig. 3-7A, gray**) could be used as a threshold to distinguish between poorly translated versus highly translated ORFs, a method established in Eukaryotes (Ingolia et al., 2011). Applying this type of analysis for newly identified ORFs from paTSS, we found that N-terminally extensions had more highly efficient translation compared to the other paTSS, even higher TE than their corresponding genes (**Fig. 3-7A, compare yellow and black**), suggesting their functional importance.

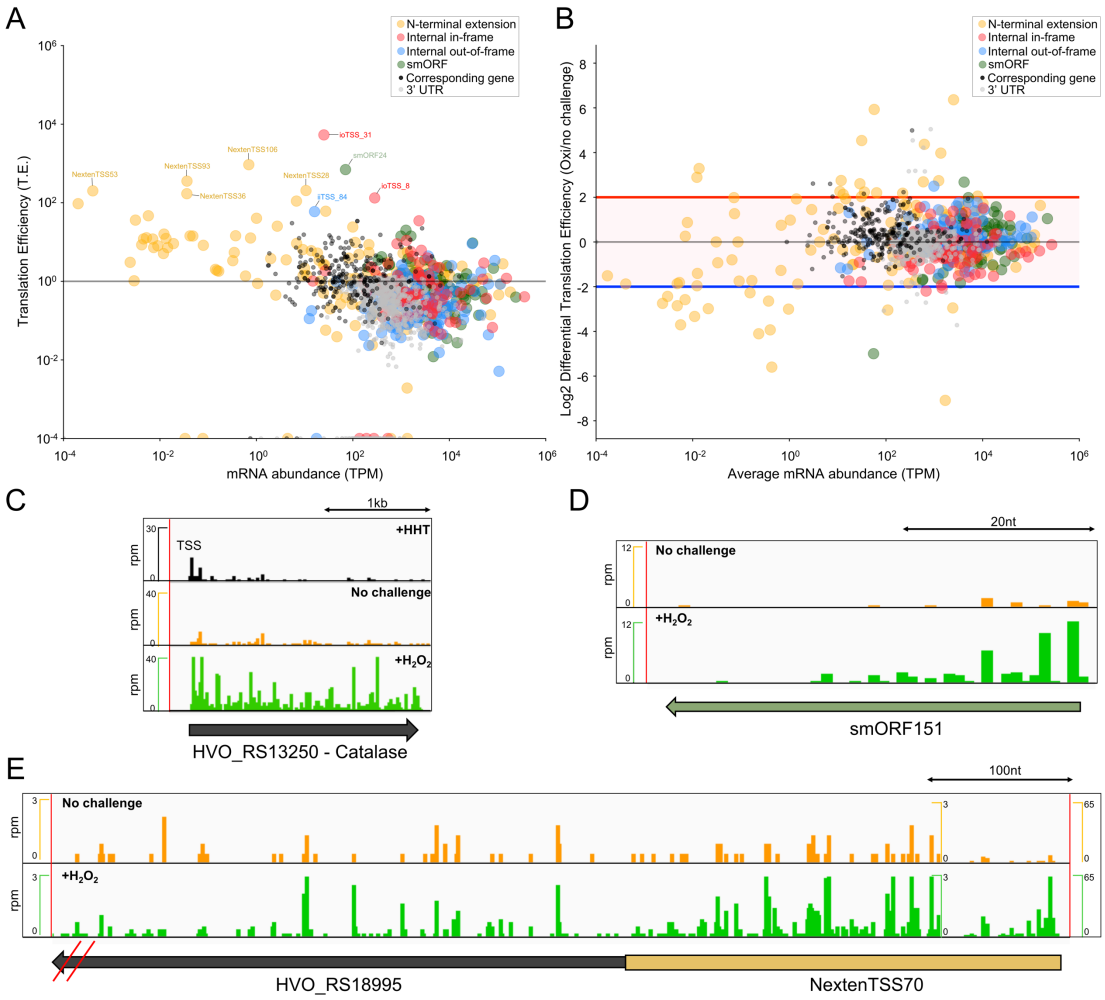

Figure 3-7: Translation of alternative TSS (aTSS). Scatterplot plot of translation efficiency of all discovered aTSS, their corresponding genes, and 3' UTRs of these genes plotted against mRNA abundance. Translation efficiency was calculated from ribosome profiling and mRNA-seq data. N-terminal extensions are in yellow, smORFs in green, internal initiation sites that are in-frame in red, and internal out-of-frame initiation sites in blue. Corresponding genes are in black and 3' UTRs are in gray, respectively. (B) MA-plot of differential translation efficiency of all aTSS during oxidative stress relative to a no challenge conditions. Colors match (A). Representative plots of differentially translated for a catalase gene (C), a smORF (D), and a N-terminal extension (E).

Changes in translation during oxidative stress

We, and others, have previously established that *H. volcanii* and other haloarchaea are highly resistant to oxidative stress (Gelsinger & DiRuggiero, 2018; Sharma, Gillum, Boyd, & Schmid, 2012). These studies have focused predominantly on physiological consequences to the stress, as well as transcriptional and post-transcriptional regulation of the response. To date, only one study has probed the protein expression profile during oxidative stress in haloarchaea using mass spectrometry (McMillan et al., 2018; Whitehead et al., 2006) and, to our knowledge, the regulation of translation during a stress response in Archaea has not been studied. Thus, we used ribosome profiling to ask how the landscape of translation changes when *H. volcanii* is exposed to the oxidative stress agent, hydrogen peroxide. Oxidative stress caused the differential translation of mRNAs similarly to what has been seen with differential expression of mRNAs in *H. volcanii* (Gelsinger & DiRuggiero, 2018) (**Fig. 3-S9**). To better assess the degree to which oxidative stress affects translation, we analyzed translation efficiency changes for all mRNAs between oxidative stress and no challenge conditions in *H. volcanii*. We found that translation efficiency for mRNAs (**Fig. 3-S10 red**) had less correlation between oxidative stress and no challenge conditions compared to the correlation between the same conditions for mRNAs translationally (**Fig. 3-S10 green**) or transcriptionally (**Fig. 3-S10 blue**). In particular, ribosome proteins, including the ribosome surveillance factor pelota, and transmembrane helix-containing proteins were up-regulated in their translation efficiency during oxidative stress. (**Fig. 3-S10 gene ontology**). This suggests that not only is oxidative stress highly regulated transcriptionally but it is also controlled at the translational level.

We next asked whether translation at aTSS might be differentially regulated during oxidative stress. We found that a subset of aTSS, predominantly enriched in N-terminally extended ORFs, were more efficiently translated (range: 4-64-fold enriched) during oxidative stress (**Fig. 3-7B**). This increase in translation could be observed on known reactive oxygen species (ROS) detoxifying genes, such as catalase (**Fig. 3-7C**), validating the observed shift in translation efficiency during stress. We found that for two specific aTSS, translation increased during oxidative stress relative to no challenge conditions; a smORF was up-regulated 5-fold (**Fig. 3-7D**) and an N-terminally extended aTSS increased 3-fold under oxidative stress (**Fig. 3-7E**). This oxidative stress-controlled N-terminally extended aTSS was upstream of a DEAD/DEAH box helicase. This helicase been implicated in stress response regulation in Bacteria by controlling mRNA decay, ribosome assembly, and translation initiation (Jiang et al., 2019), which was in accordance with our previous observation that ribosome proteins and surveillance factors were differentially regulated translationally (**Fig. 3-S10 gene ontology**). Overall, translational regulation appears to be an important mechanism for *H. volcanii* , and other haloarchaea, in their stress response and may be involved in their increased resistance to oxidative stress.

DISCUSSION

Translation is a highly regulated process and it represents the single largest investment of energy in the cell (Russell & Cook, 1995; Verduyn et al., 1991). Studies of the mechanisms and regulation of protein synthesis in bacteria and eukarya have been greatly facilitated by the development of an experimental approach to globally analyze

the full set of ribosomes engaged in translation, a high-throughput technique termed ribosome profiling. In comparison, due to a lack of comparable tools, we know relatively little about translation in Archaea (Brenneis & Soppa, 2009; Ingolia, et al., 2019). Here we developed a ribosome profiling method in *H. volcanii* and revealed the first global view of translation in an archaeon and in an extremophile. In the future, we anticipate that ribosome profiling will provide further insights into translational control of gene expression in *H. volcanii* and that this protocol will be adapted to other members of the Archaea.

A critical challenge in the development of a robust and reproducible ribosome profiling protocol for a new organism is ensuring that translation is arrested in a way that faithfully captures the *in vivo* translational landscape. This problem is particularly acute when working with extremophiles because of the harsh experimental conditions required to maintain the integrity of their macromolecules; harvesting the culture places a strain on cells that can quickly and pervasively alter translation levels. In yeast, the preferred method is to harvest cultures by filtration followed by flash freezing; we were unable to obtain intact ribosomes from filtered *H. volcanii* samples. Pre-treating the culture with a translation elongation inhibitor (anisomycin, ANS) inhibited translation well enough that we were able to obtain 70S monosomes from sucrose gradients after harvesting cultures by centrifugation. However, as reported for other microorganisms, pre-treatment with antibiotics led to problems with reproducibility, at least partially due to accumulation of ribosomes at the 5'-end of transcripts (given that initiation continues as elongation is arrested) (Gerashchenko & Gladyshev, 2014; Mohammad et al., 2019). To alleviate these issues, we used a direct flash-freezing method that allowed us to faithfully arrest

translation after cell lysis and preserve the *in vivo* translational landscape. In this method, first reported for *E. coli* (Mohammad et al., 2019), cultures were directly flash-frozen with no antibiotics prior to cell lysis.

A second challenge in biochemical studies of protein synthesis in *H. volcanii* is that the proteome and the ribosomes require high salt concentrations to maintain correct folding and intermolecular interactions (Gunde-Cimerman et al., 2018). Previous studies in *H. volcanii* have been unable to isolate intact 70S monosomes, probably due to subunit dissociation *in vitro* following cell lysis (Kramer et al., 2014; Wyss et al., 2018). We were successful in developing lysis conditions and buffers that allow us to isolate 70S monosomes, a crucial step for ribosome profiling. Even with our optimized lysis buffer, however, we observed that monosome and polysome abundance declines over time after cell lysis, suggesting that ribosome subunits were dissociating with time. More optimization to resemble *H. volcanii* intracellular salt concentrations (i.e. diversity of salts) may help. The challenge of ribosome instability *in vitro* will likely also affect attempts to perform ribosome profiling in other extremophiles such as hyperthermophiles and anaerobes, where environmental conditions are temperature and oxygen-limited. It may be that crosslinking strategies can stabilize 70S ribosomes on mRNA transcripts to overcome these challenges, although in our hands crosslinking did not noticeably alter monosome/polysome levels in *H. volcanii* (data not shown).

Although we isolated a broad range of footprints (10 – 45 nt), we found that the major footprint from elongating ribosomes in *H. volcanii* is 27 nt long. Because these footprints are trimmed back to the edges of the ribosome, they give the most precise information about the position of the ribosome. The fact that > 75% of 27 nt footprints

map to the first position of codons indicates that ribosome profiling in *H. volcanii* can capture reading frame at codon resolution, unlike studies in bacteria (Gerashchenko & Gladyshev, 2014; Mohammad et al., 2019). The size of the predominant *H. volcanii* footprint is close to the size of the major eukaryotic footprint (28 nt in yeast), another example of the close evolutionary relationship between Archaea and Eukarya. This is surprising given that archaea have smaller ribosome subunits, more similar in size to bacterial than eukaryal ribosomes. However, the relatively higher number of translation factors found in archaea, most of which are homologous to eukaryotic translation factors, could conceivably produce a larger footprint (Bell & Jackson, 1998). Alternatively, this larger footprint size may be related to the way rRNA pair with mRNA upstream of the P-site codon in *H. volcanii*. Going forward it will be interesting to further characterize protein synthesis at the genome-wide level in other members of the third domain of life to shed light on this archaeal-eukaryal evolutionary relationship. In particular, studying translation and major footprint(s) of Asgard archaea, the closest evolutionary relative of the eukaryotic nucleus, will undoubtedly help towards this goal (Sprang et al., 2015).

In addition to the predominant 27 nt footprints from elongating ribosomes, we also observed a distribution of shorter footprints (< 20 nt). In the context of a primarily leaderless transcriptome (>70% in *H. volcanii*), ribosomes that initiate on a leaderless mRNA will have an mRNA channel that is empty upstream of the P site. After nuclease digestion, mRNA footprints from ribosomes on leaderless start codons will thus produce a smaller footprint size (peak at 16 nt) compared to an elongating ribosome (27 nt) that has the entire tunnel occupied by transcript. In contrast, ribosomes on start codons on leadered mRNAs have a longer footprint size due to the 5' UTR sequence that occupies

the mRNA channel during initiation. Indeed, we observed an enrichment of 27 nt footprints (enhanced with elongation inhibitors) upstream of start codons on the mRNAs that are leadered in *H. volcanii.* This is of particular interest because our understanding of leaderless initiation, thought to be the evolutionarily oldest mechanism, is still poorly characterized (Andreev et al., 2006; Brenneis & Soppa, 2009; Dennis, 1997). Our ability to distinguish between footprint sizes of leaderless and leadered transcripts, therefore, provides a model to address outstanding questions regarding the mechanisms between these different forms of initiation.

To our knowledge only two other studies have investigated the differences between leadered and leaderless mRNAs using ribosome profiling (Jeong et al., 2016; Shell et al., 2015). Neither reported differences in footprint lengths at start codons on leadered and leaderless transcripts but notable differences with their work and ours are that (i) the Bacteria studied did not have predominantly leaderless transcriptomes (*Mycobacterium smegmatis* 20%, *Streptomyces coelicolor* 21%), and (ii) a limited range of footprints sizes were size selected (*Mycobacterium smegmatis* 28 nt, *S. coelicolor* 26-32 nt) which may have prevented such analysis (Jeong et al., 2016; Shell et al., 2015). In the future, studies should isolate a broad distribution of footprint sizes to investigate how general footprint size differences are for leadered versus leaderless mRNAs across the three domains of life.

One of the strengths of ribosome profiling is its ability to detect differences in local elongation rate that occur as ribosomes pause during elongation. Ribosome pausing due to non-optimal codon usage or by interactions between the nascent peptide and the ribosomal exit tunnel can regulate gene expression (Buskirk & Green, 2017). We used a

targeted drug approach, treating *H. volcanii* with serine hydroxamate (SHX), to starve the cells of a specific aminoacyl-tRNA and detect pauses at the level of codons. To our surprise, SHX did not yield pauses as Ser codons, as it does in bacteria, but instead caused strong pausing at Met codons. By measuring charged and uncharged tRNAs, we found evidence that SHX blocks the charging of Met-tRNA, either directly or by inhibiting Met biosynthesis (a serine-dependent process).

We further assessed the pausing landscape in *H. volcanii*, using untreated cells, and found reproducible Pro pauses at A-, P-, and E-sites. Pro is known to be both a poor peptidyl acceptor and donor and polyproline sequences are particularly problematic. A specialized elongation factor (EFP in bacteria and eIF5A in eukaryotes) helps to mitigate these pauses by accelerating peptidyl transfer between Pro residues (Gutierrez et al., 2013). Our data suggest that polyproline pausing also occurs in *H. volcanii*, which encode a related elongation factor, aIF5A (Wagner & Klug, 2007). Many of these pause effects were masked by ANS treatment, prior to harvesting (either flash freezing or centrifugation), likely because the drug inhibits elongation with some level of sequence selectivity. The observation of pauses reflecting known limitations in translation (e.g. Pro), and the disruption of these pauses by translation inhibitors, strongly suggests we are capturing true snapshots of *in vivo* biology that is not masked by methodology-based artifacts. Our results confirm that ribosome pausing occurs in Archaea with implications for gene regulation, an observation already demonstrated in the other domains of life (Buskirk & Green, 2017; Collart & Weiss, 2019; Darnell et al., 2018; Gong & Yanofsky, 2002; Karlsen et al., 2018; Nakatogawa & Ito, 2002; Yanofsky, 1981).

Mapping translation start sites (TSS) transcriptome-wide with antibiotics to trap initiation complexes has proven to be a powerful strategy in eukaryotes and bacteria (Ingolia et al., 2011; Meydan et al., 2019; Weaver et al., 2019). The need for accurate identification of TSS is particularly apparent in Archaea where most gene annotations are generated from general computational pipelines that are not totally reliable. Further complicating bioinformatics analyses, most genes identified in Archaea are not well conserved with known genes in other organisms. Using harringtonine (HHX) to lock ribosomes onto initiation sites, we accurately identified 51% of the TSS for annotated genes from the NCBI RefSeq annotation. We also identified potential protein-coding sequences in contexts that are often missed (Jevtić et al., 2019), such as TSS on extremely small ORFs or those that are within ORFs or antisense to known ORFs. We found hundreds of putative alternative TSS in *H. volcanii* including initiation sites upstream of annotated TSS (N-terminal extensions), initiation sites within annotated ORFs that could produce truncated proteins (internal in-frame) or completely new proteins (internal out-of-frame), and small ORFs (<50 amino acids). These putative alternative proteins may have been previously obscured in proteomic data but may play important roles in cell physiology and stress response. N-terminally extended and internal initiation could play important roles in gene regulation of their corresponding ORFs by altering translation efficiency of the annotated protein, as proposed in Bacteria and Eukarya (Ingolia et al., 2011; Meydan et al., 2019). The verification of these putative alternative proteins is challenging in Archaea, compared to Bacteria and Eukarya, because genetic tools are less developed and the limitations on biochemical approaches due to proteomic adaptations to extreme environments (Kubatova et al., 2019).

CONCLUSION

In conclusion, we coupled ribosome profiling with translation inhibitors to determine essential characteristics of translation in a member of the Archaea, for the first time. Specifically, we (i) determined the size of the archaeal ribosome footprint, (ii) assigned translation states of the ribosome to footprint lengths in a majority leaderless transcriptome, (iii) experimentally induced ribosome pauses and clarified the pausing landscape comprehensively, (iv) identified novel proteins including small open reading frames (smORFs), and (v) provided evidence that many genes initiate at putative alternative translation start sites (aTSS) around and within open reading frames (ORFs). This work demonstrates how a microorganism with a gene-dense genome can potentially produce proteins with distinct functions (isoforms) using the same gene. Lastly, ribosome profiling revealed the features archaea use in their translational apparatus, which are both a mosaic of bacteria and eukarya as well as features unique to their domain.

MATERIAL AND METHODS

H. volcanii culture conditions, harvesting, and cell lysis

H. volcanii H98 single colonies were picked and grown overnight at 42°C with shaking at 220 rpm (Amerex Gyromax 737) in Hv-YPC medium supplemented with thymidine (50 μg/mL final concentration) (Dyall-Smith, 2009) until saturation (OD_{600} > 1.0). These cultures were then diluted to OD_{600} 0.02 in fresh media, grown to OD_{600} 0.4 and split evenly into two flasks; one flask was used as a no-treatment control and the other was treated with either 1 mg/mL homo-harringtonine (HHT, Sigma Catalog #SML1091), 20 mM serine hydroxamate (SHX, Sigma Catalog #S4503), or 100 μg/mL anisomycin (ANS, Sigma Catalog #A9789) with final concentrations as indicated in the

text. Cells were harvested by centrifugation, filtering, or direct freezing of the culture in liquid nitrogen.

For cells harvested by centrifugation, cultures were immediately centrifuged at 8,600 x g for 3 min at room temperature (RT), the supernatant removed, and the pellets flashed frozen in liquid nitrogen. For cell lysis, the frozen pellets were resuspended in 1 mL of 1x lysis buffer (3.4 M KCl, 500 mM $MgCl_2$, 50 mM $CaCl_2$, 1 M Tris pH 7.5) with an additional 100 μg/mL ANS, transferred to a cryo-mill (Spex SamplePrep 6870), and pulverized with 5 cycles (1 min grinding at 5 Hz, 1 min cooling). Lysates were then thawed at RT, centrifuged at 10,000 x g for 5 min, and transferred to a new tube on ice. For cells harvested by filtration, cultures were immediately poured into a filtration unit (90 mm, Sigma WHA1950009) on a 0.25 μm filter. As soon as cells began to accumulate on the filter (for < 1 min), cells were collected using a sterile spatula, flash-frozen in liquid nitrogen, and lysed in a cryo-mill in the same manner as the centrifugation harvesting method.

For direct freezing of cultures in liquid nitrogen, 100 mL of culture were sprayed directly into liquid nitrogen using a serological pipette. The frozen culture formed small pellets that were collected, and 50 g of pellets were weighed to add 1x lysis buffer (3.4 M KCl, 500 mM $MgCl_2$, 50 mM $CaCl_2$, 1 M Tris pH 7.5) and 100 μg/mL ANS to prevent ribosome elongation when thawed later on. The pellets were then pulverized in a cryo-mill for 10 cycles due to the larger volume of input (1 min grinding at 10 Hz, 1 min cooling). The lysates were thawed at RT and ribosomes were pelleted over a 60% sucrose cushion (sucrose dissolved in lysis buffer) in an ultracentrifuge with a Ti-70 rotor at

26,4902 x g (60,000 rpm) for 2 hrs at 4°C. Ribosome pellets were resuspended in 200 µL lysis buffer.

Determination of translation inhibitor concentrations

The following translation inhibitors were tested in *H. volcanii*: ANS, cycloheximide (Sigma Catalog #C1988), HHT, SHX, thiostrepton (Sigma Catalog #T8902), and tetracycline (Sigma Catalog #T8032). These inhibitors were tested either to prevent ribosome elongation after cell lysis or to alter translation in order to validate our ribosome profiling method. Serial dilutions by one order of magnitude of each drug (ANS 1 µg/mL, 10 µg/mL, 100 µg/mL; cycloheximide: 10 µg/mL, 100 µg/mL, 1 mg/mL; HHT: 500 µg/mL, 1 mg/mL; SHX: 0.2 mM, 2 mM, 20 mM; thiostrepton: 5 µg/mL, 50 µg/mL, 500 µg/mL, tetracycline: 5 µg/mL, 50 µg/mL, 500 µg/mL) were administered to *H. volcanii* liquid cultures and incubated, as described above. 100 µL aliquots were removed to measure the optical density of the culture (600 nm) over a 24-48hrs time-period. Concentrations of drugs that completely halted growth were used as the final concentrations in all ribosome profiling libraries.

Sucrose gradients for ribosomes and subunits

Cell lysates (from cells harvested by centrifugation or direct freezing) were loaded onto a 10-50% sucrose gradient in 1x lysis buffer. Gradients were centrifuged in an ultracentrifuge using a SW41 rotor at 90,140 x g (35,000 rpm) for 2.5 hrs at 4°C. Gradients were then fractionated into 400 µL fractions to resolve ribosome 30S, 50S subunits, monosomes, and polysome fractions. Fractions were flash frozen on dry ice for later RNA isolation.

RNA extractions and sucrose fraction characterization

RNA was isolated from sucrose gradient fractions that corresponded to ribosome subunits, monosomes, and polysomes. Briefly, 250 µL Trizol LS (Lifesciences) was added to gradient fractions, vortexed, and incubated at RT for 1 min. 150 µL chloroform was added to the samples, vortexed, and centrifuged at 14,000 x g for 10 min at 4 °C. The mixtures were separated into a lower red phenol-chloroform, an interphase, and a colorless upper aqueous phase that was transferred into a new tube. Glycoblue (NEB) and an equal volume of 2-propanol were added to the samples followed by an incubation of 30 min on dry ice. Samples were then centrifuged at 14,000 x g for 30 min at 4 °C to pellet the RNA. RNA pellets were washed twice with 75% ethanol, air dried for 2 min, and resuspended in nuclease-free water. To determine whether sucrose gradient peaks corresponded to 30S, 50S, monosomes, or polysomes, RNAs extracted from each fraction were separated on a denaturing agarose gel. 16S and 23S rRNA bands were visualized by SYBR Gold staining (ThermoFisher).

Optimization of MNase activity in high salt buffer

MNase (Nuclease S7, Roche Cat: 10107921001) activity was tested at increasing KCl, $MgCl_2$, and $CaCl_2$ concentrations in lysis buffer in a 2-step double stranded DNA (dsDNA) absorbance assay. In the first step, serial dilutions of MNase were added to a substrate buffer (2 g/mL salmon sperm dsDNA, 10 mM Tris pH 7.5, lysis buffer) in 96-well plates, shaken for 10 s, and OD at 260 nm was measured every min for 1 h at 25°C. In a second step, the same assay was performed at the optimal concentration of MNase with increasing concentrations of $CaCl_2$ in the lysis buffer.

Footprinting of ribosomes and Ribosome profiling library preparation

Cell lysates (from cells harvested by centrifugation or direct freezing) were processed by first treating 20 absorbance units (AU, Nanodrop) of lysate RNA with 12,000 units of MNase for 1 hr at 25°C. After nuclease digestion, samples were loaded onto 10-50% sucrose gradients and RNA was isolated from monosome fractions as described above. Library preparation was performed as previously described (Mohammad et al., 2019). Briefly, 10 µg of RNA fragments were used to purify 10-45 nt RNA fragments by gel electrophoresis (PAGE 15% TBE 7 M Urea gel); RNA fragments were treated with T4 polynucleotide kinase (NEB), ligated to the linker (NEB Universal miRNA Cloning Linker) using T4 RNA ligase (NEB), and gel extracted from a 10% TBE Urea gel. rRNAs were subtracted from the RNA fragments using the Ribo-Zero rRNA removal kit for bacteria (Illumina). The resulting fragments were reverse transcribed with SuperScript III (Invitrogen) using custom primers previously described (Mohammad et al., 2019). Template RNA was degraded using NaOH for 20 min at 98 °C, the fragment were gel extracted from a 10% TBE Urea gel, circularized using CircLigase (Epicentre) and PCR amplified (8-12 cycles) with Phusion polymerase (NEB) using custom primers (Mohammad et al., 2019). PCR products were gel extracted from a 10% TBE gels and analyzed for size and concentration using a BioAnalyzer high sensitivity DNA kit (standard protocol) before sequencing on an Illumina HiSeq 2500 at the Johns Hopkins Genetic Resources Core Facility (Baltimore, MD).

Transcriptome reannotation

We used RNA-seq data we previously published (Gelsinger & DiRuggiero, 2018) and the program Rockhopper2 (default settings for paired-end reads) (Tjaden, 2015) to accurately determine coordinates for all transcription start sites in the *H. volcanii*

transcriptome. Using the output of Rockhopper2, an annotation file (.gff) was created to account for untranslated regions (UTR) both 5' and 3' with respect to previously annotated translation start sites (TSS) as well as operonic transcripts. All subsequent ribosome profiling analysis used these transcriptome reannotations instead of database deposited annotations (e.g., NCBI and UCSC).

Ribosome profiling data analysis

All ribosome profiling data were analyzed using previously established methods and python scripts in *E. coli* (Mohammad et al., 2019). In brief, reads were trimmed using trim_galore, reads corresponding to rRNA and tRNA were discarded, and the remaining reads were mapped against the *H. volcanii* NCBI RefSeq genome (taxonomy identification [taxid] 2246), allowing two mismatches using Bowtie v 0.12.7 (Langmead et al., 2009). These alignments were then used in custom python scripts to: (a) calculate read density (assigned to the 3'-end of reads) across the genome, (b) calculate average ribosome position and read length distributions along ORFs using meta-gene analysis (3'- or 5'-end of reads as noted) on all transcripts, leadered transcripts, or leaderless transcripts, (c) calculate average amino acid and codon pause scores, (d) asymmetry scores for ORFs, and (e) expression for each gene (Mohammad et al., 2019). Custom bash/awk scripts were used to separate 17 nt and 27 nt footprint reads and these alignments were analyzed in the same manner.

TSS identification with harringtonine (HHT) ribosome profiling libraries was done using previously published custom python scripts that were modified for our purposes (Meydan et al., 2019; Weaver et al., 2019). Identification of novel TSS and smORFs was done as previously published (Meydan et al., 2019; Weaver et al., 2019)

with the following changes: (a) rpkm was calculated for each TSS peak in both the HHT and no challenge library, (b) the TSS peak rpkm were >5-fold enriched in HHT compared to no challenge, and (c) the ORF rpkm were >5-fold enriched in no challenge compared to HHT. Identification of known TSS, internal TSS (iTSS), and N-terminal extensions was done as previously published using density files assigned to the 3'-end of reads and adjusted for the P-site (15 nt offset) (Meydan et al., 2019).

Translation efficiency analysis was done using previously published RNA-seq data (Gelsinger & DiRuggiero, 2018). Thresholding of expressed CDS were done using custom python scripts, and transcript per million (TPM) was calculated for each CDS and aTSS using ribosome profiling and corresponding RNA-seq samples. Translation efficiency was calculated by dividing the ribosome profiling TPM by the RNA-seq TPM per CDS and per aTSS.

tRNA northern blot of tRNA charging

To determine tRNA charging of serine and methionine tRNAs, a tRNA deacylation and β-elimination treatment was performed as previously described (Mohammad et al., 2019). Briefly, total RNA was extracted from no treatment and SHX-treated flash frozen cells using the Quick RNA extraction kit (Zymo) following the standard protocol. The RNA was then DNase I (NEB) treated for two hrs using 2 units of DNase I per hour and was cleaned of residual DNase I with the RNA clean and concentrator-5 (Zymo). DNase-treated RNA was then divided into two equal aliquots; one of the aliquots was deacylated by treatment with 1M Tris pH 9 at 37 °C for 1 hr, and then ethanol precipitated. Following deacylation, both aliquots were treated with sodium periodate and 1M lysine to promote β-elimination of oxidized 3' RNA ends, and then ethanol precipitated.

Samples were then run on a 10% TBE 7 M Urea denaturing polyacrylamide gel. RNA was transferred using a wet transfer apparatus (Hoefer TE62) onto a nylon membrane and UV crosslinked to the membrane using the automatic setting (UV Stratalinker 1800). Membranes were probed in Ultrahyb Oligo buffer (Ambion) with 5'-32P-labeled ($tRNA^{Ser}$) TCACGTGTCCGAATGGACAGTAGA or 5'-32P-labeled ($tRNA^{Met}$) ATGAGCCCGGCGGAATCTCCT and signal was detected on a Typhoon phosphoimager.

Data availability

Ribosome profiling reads, density bigwigs, transcriptome reannotation, and translation start site annotations are available on NCBI GEO: GSE138990. Data analysis scripts are available on github: https://github.com/dgelsin/Ribosome_profiling_MS

CONTRIBUTIONS

DRG: Conceptualization, Investigation, Methodology, Project administration, Writing – original draft, Writing – review & editing

ED: Investigation – SHX tRNA charging northern blots

RR: Investigation – inhibitor growth curves

FM: Methodology, Conceptualization

AB: Conceptualization, Funding acquisition, Project administration, Supervision, Validation, Writing – review & editing

JDR: Conceptualization, Funding acquisition, Project administration, Supervision, Validation, Writing – review & editing

ACKNOWLEDGEMENTS

We thank Dr. Rachel Green and her laboratory for reagents and helpful discussions. We thank Dr. Colin Wu for assistance in analyzing ribosome profiling data for Eukaryotes. This work was supported by the National Aeronautics and Space Administration [18-EXO18-0091 to J.D.], the Air Force Office of Scientific Research [FA9950-14-1-0118 to J.D.], and the National Institutes of Health [GM110113 to A.B.].

SUPPLEMENTAL FIGURES

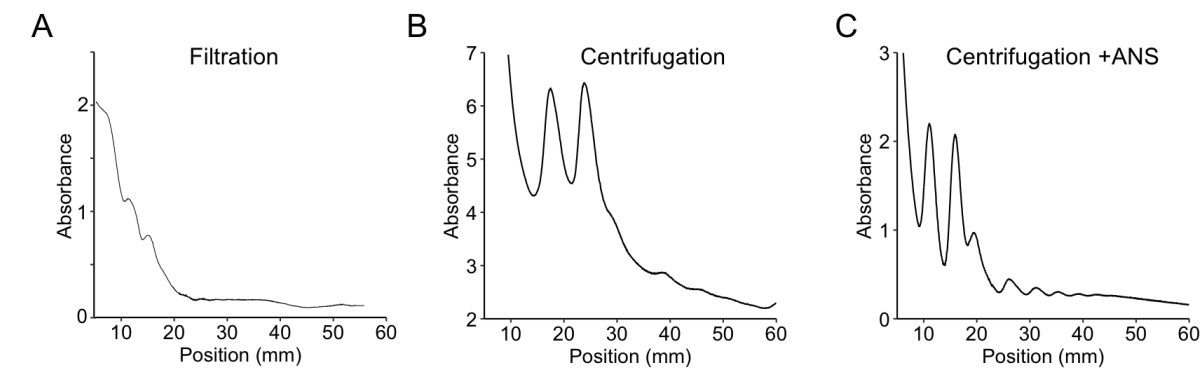

Figure 3-S1: Sucrose gradient plots of (A) cells filtered and harvested with liquid nitrogen, (B) cells harvested with centrifugation and no-drug added, and (C) cells harvested with centrifugation with anisomycin (ANS) added.

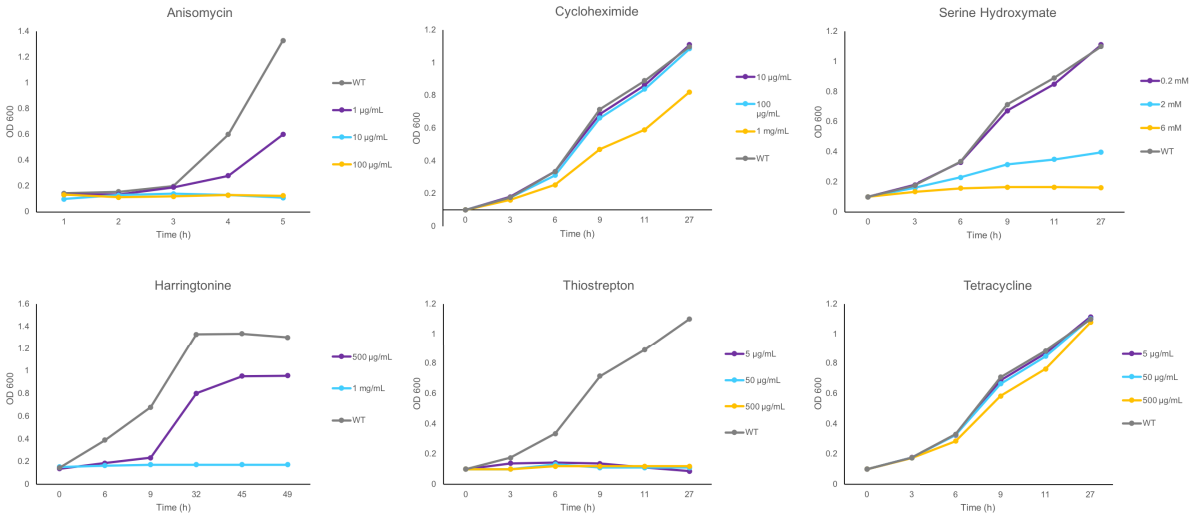

Figure 3-S2: Growth curves of *H. volcanii* exposed to translation inhibitors.

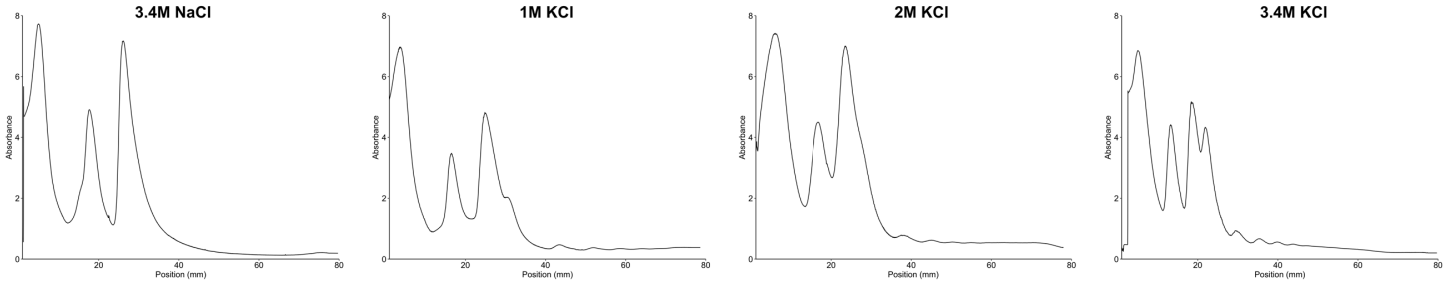

Figure 3-S3: Dependency of ribosome integrity in *H. volcanii* on potassium chloride (KCl) levels shown with sucrose gradient plots.

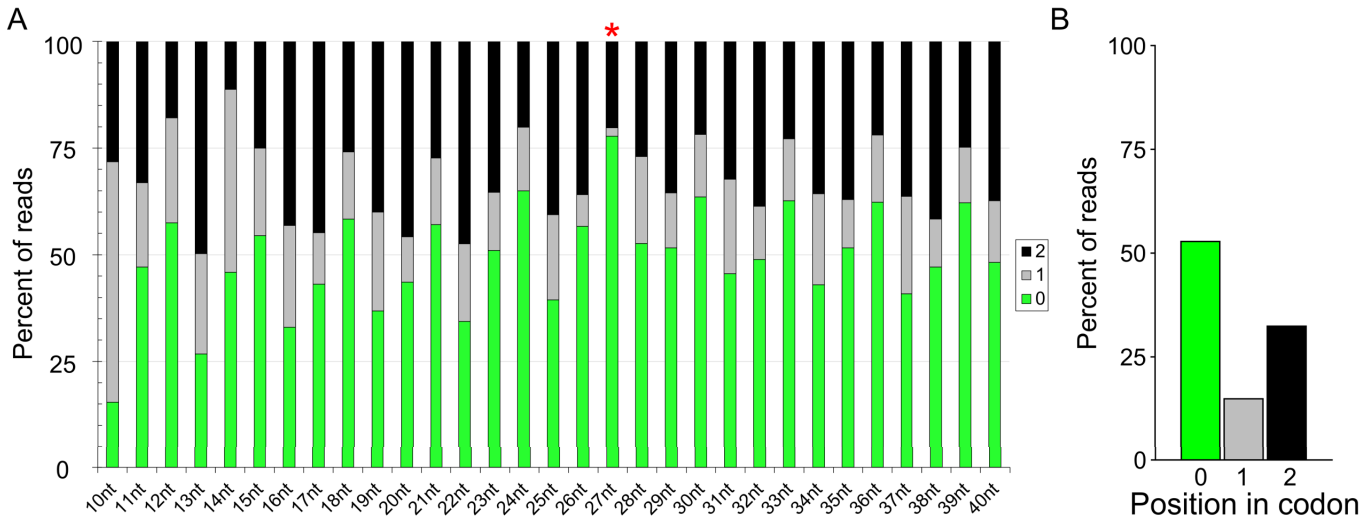

Figure 3-S4: Mapped distribution of the elongating footprints to the position in codon (0, +1, or +2) for (A) each individual footprint ranging from 10-40 nt, and (B) the average of all footprints.

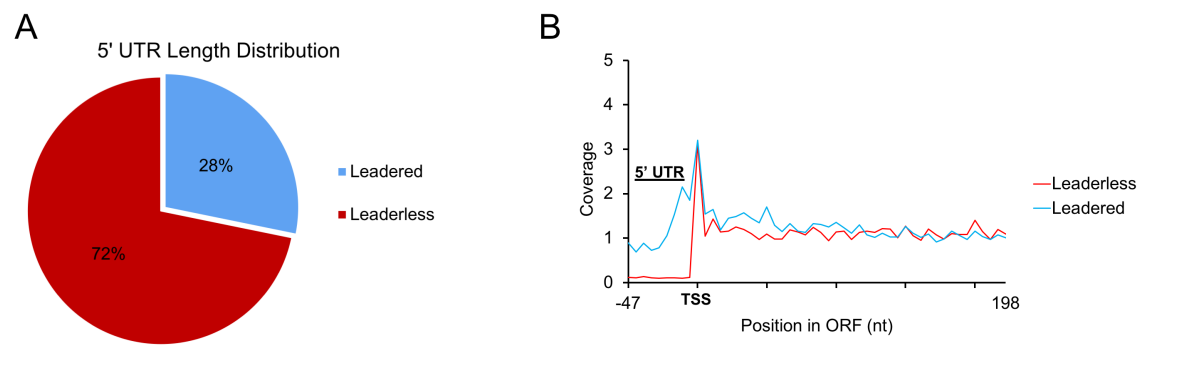

Figure 3-S5: Experimental identification of leadered and leaderless mRNAs in *H. volcanii.* (A) 5' UTR distribution of the transcriptome in *H. volcanii.* (B) Meta-gene analysis plot of RNA-seq reads on leadered and leaderless transcripts.

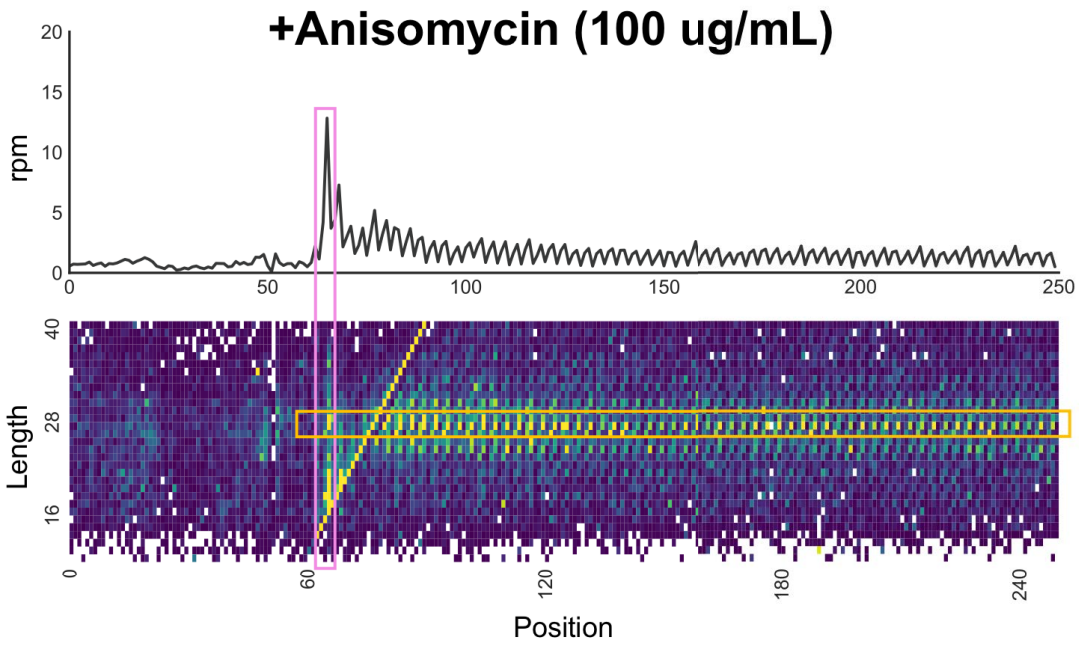

Figure 3-S6: 3'-mapped meta-gene analysis at start codons of all footprints (10-40 nt) for anisomycin (ANS)-treated cells. Short footprints are boxed in pink and 27 nt footprints are boxed in yellow, corresponding to Fig 3 C and D.

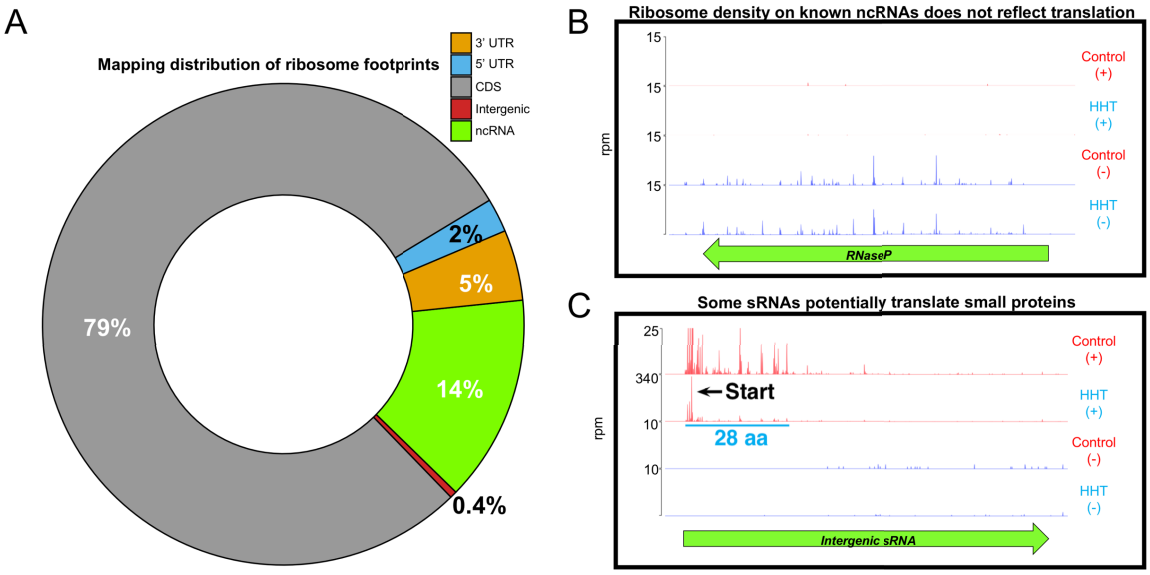

Figure 3-S7: Footprint mapping distribution to different transcriptome features. (A) Pie chart of ribosome footprints that mapped to the *H. volcanii* transcriptome, sectioned by transcriptional elements. Representative examples of non-coding RNAs with ribosome density along the transcript with and without harringtonine (HHT) treatment: (B) a non-coding RNA (RNase P) with no altered ribosome density and (C) an intergenic sRNA with altered ribosome density due to HHT treatment. Red bars are no drug treatment ribosome profiling libraries (control) and blue are HHT ribosome profiling libraries. Plus strand (+) and minus strand (-) are plotted separately.

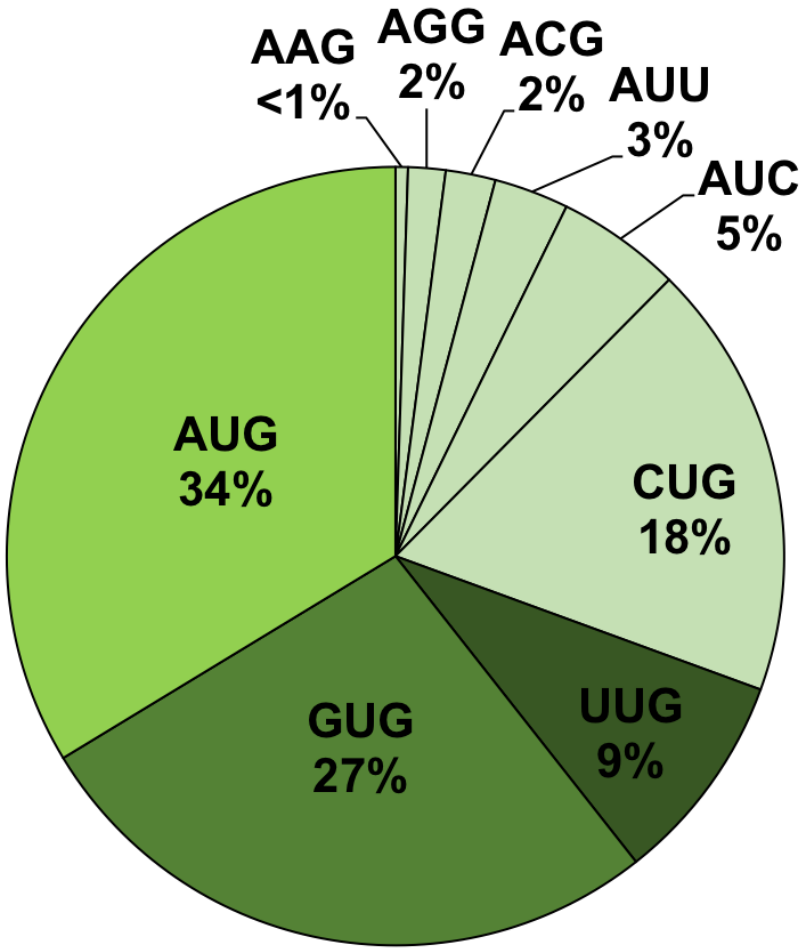

Figure 3-S8: Distribution of all codons detected at initiation sites for HHT-identified paTSS.

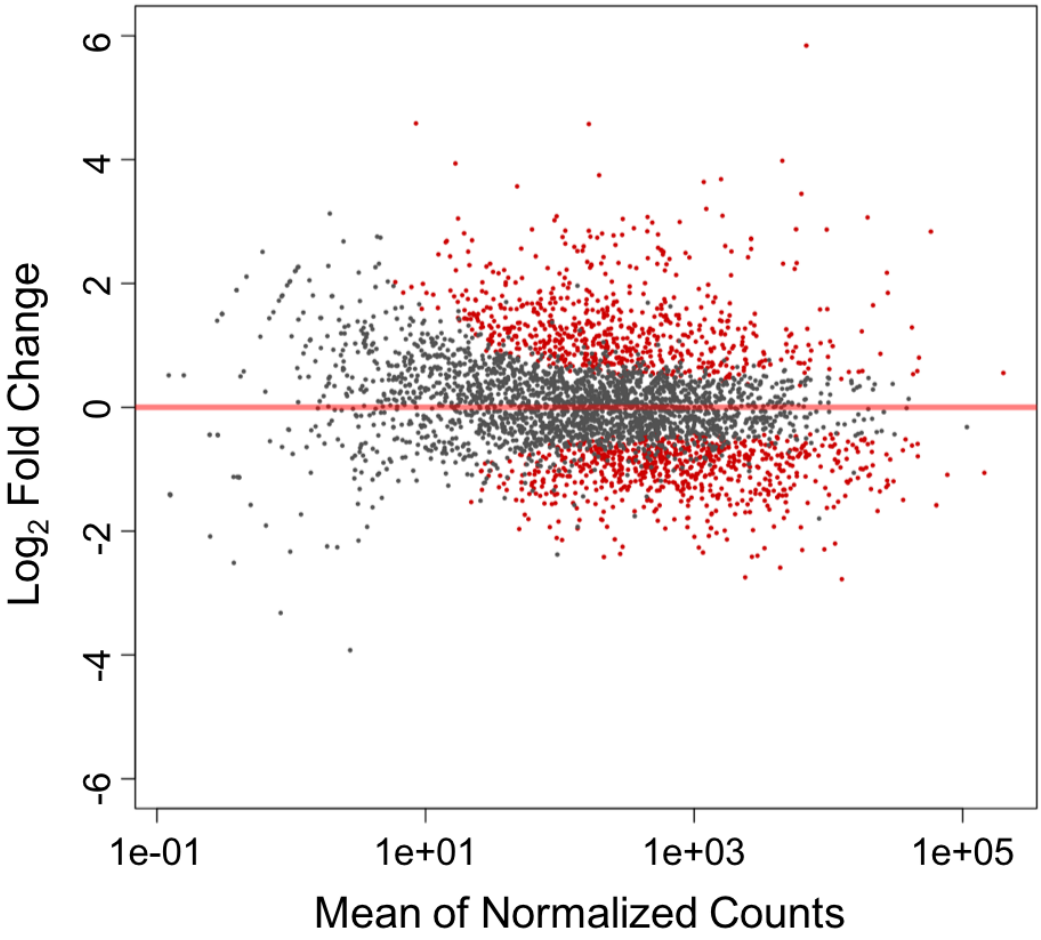

Figure 3-S9: MA-plot of differentially translated mRNAs during oxidative stress (2 mM H_2O_2) versus no challenge conditions in WT *H. volcanii*

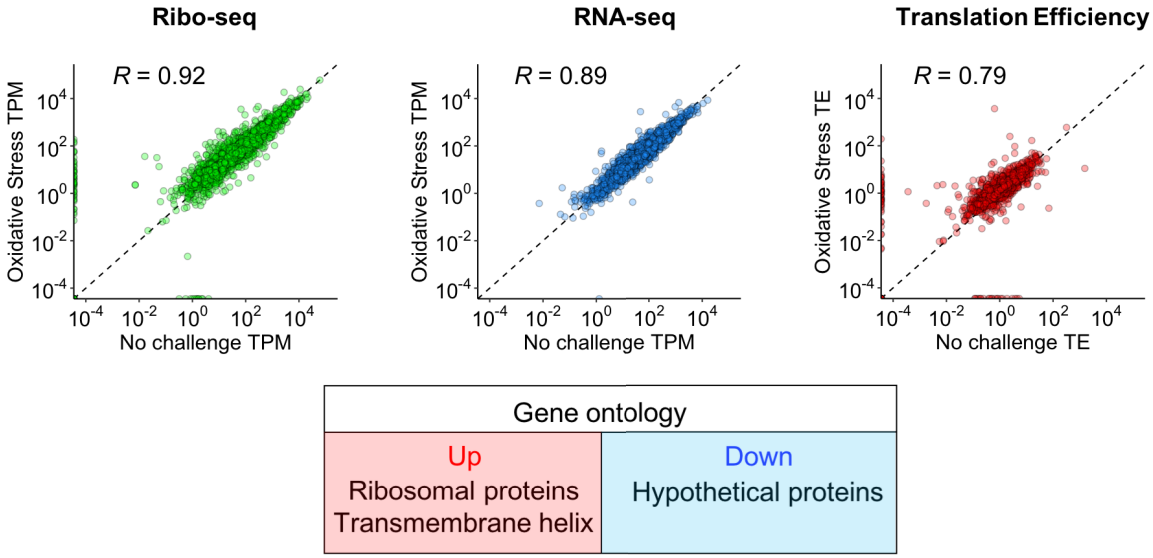

Figure 3-S10: Correlation plots of translation, transcription, and translation efficiency (TE) between oxidative stress and no challenge conditions. Gene ontologies of significant differential TE genes are provided.

Chapter 4

Post-transcriptional regulation of redox homeostasis during oxidative stress in haloarchaea by the small RNA SHOxi

To be submitted to RNA in 2020 – Gelsinger DR, Reddy R, Whittington K, Debic S, & DiRuggiero J, RNA 2020.

ABSTRACT

Previous work has shown that *Haloarchaea* are an order of magnitude more resistant to oxidative stress than most mesophilic organisms. Despite this resistance, a comprehensive understanding of the mechanisms regulating this remarkable oxidative stress response is lacking. Recently, small non-coding RNAs (sRNAs) were discovered in Archaea under a variety of environmental conditions, suggesting that, as in Bacteria, stress response pathways might be regulated by sRNAs. Previously we discovered sRNAs that were differentially expressed in response to oxidative stress in the archaeon *Haloferax volcanii*. A knockout strategy was used to elucidate the functional role of the most up-regulated intergenic sRNA named **S**mall RNA in **H**aloferax **Oxi**dative Stress (**SHOxi**), an 87% GC transcript. *ΔSHOxi* survival to an acute dose of H_2O_2 was significantly lower than that of the wild type. SHOxi was predicted to form a stable secondary structure with a 30-nucleotide conserved loop region as the potential binding site for *trans*-targets. *In vivo* targets of SHOxi were identified using a knockout and

overexpression RNA-seq strategy, with 46 potential targets that changed expression in *ΔSHOxi* + H_2O_2. NAD-dependent malic enzyme mRNA was the only transcript that passed both the *in silico* and *in vivo* screens and had its stability decreased in the presence of SHOxi during oxidative stress. We measured NADH levels and found that they were lower in the presence of SHOxi. Measuring protein carbonyls in *ΔSHOxi* compared to the wild type under oxidative stress demonstrated that more reactive oxygen species (ROS)-damage occurred in the absence of SHOxi. We tested SHOxi-Malic enzyme mRNA base-pairing interactions by site directed mutagenesis in an overexpression vector and found that mutating the predicted binding site increased malic enzyme mRNA levels. SHOxi is thus a regulator of redox homeostasis during oxidative stress in *H. volcanii* through post-transcriptional destabilization of malice enzyme mRNA, which is instrumental to the survival of this archaeon in the presence of extreme ROS levels. This study is the first to establish the regulatory effects of sRNAs on mRNAs during the oxidative stress response in Archaea.

INTRODUCTION

Small non-coding RNAs (sRNAs) have been established as important regulators for multiple cellular functions across the 3 domains of life (Cech & Steitz, 2014). sRNAs are ubiquitous in Bacteria and Eukarya, playing essential roles in transcriptional regulation, RNA processing and modification, mRNA stability, and translation regulation (Cech & Steitz, 2014; Storz et al., 2011; Wagner & Romby, 2015). More recently, sRNAs have been discovered in Archaea and found to be particularly abundant in the haloarchaea (Babski et al., 2016; Gelsinger & DiRuggiero, 2018b, 2018a; Laass et al., 2019; Wyss et al., 2018). To date, very few of these newly reported candidate sRNAs have been functionally characterized (Buddeweg et al., 2017; Kliemt et al., 2019; Prasse et al., 2017), and many questions remain.

Archaeal sRNAs range from 50 to 500 nucleotides in size and can be categorized into three classes: (1) intergenic sRNAs, (2) antisense sRNAs, and sense sRNAs (Babski et al., 2016; Gelsinger & DiRuggiero, 2018b; Laass et al., 2019). RNA sequencing (RNA-seq) applied to a limited number of archaeal species, including *Haloferax volcanii, Methanosarcina mazei,* and *Sulfolobus solfataricus*, revealed that hundreds to thousands of sRNAs were potentially transcribed from those gene dense genomes. Some of these sRNAs have been implicated in several biological functions such as cellular growth, osmolarity, carbon and energy metabolism, nutrient uptake, stress response, and biofilm formation, which underscores their importance for cellular functionality (Gelsinger & DiRuggiero, 2018b; Jaschinski et al., 2014; Kliemt et al., 2019; Orell et al., 2018; Prasse et al., 2017). For example, in the methanogen *M. mazei*, $sRNA_{154}$ was up-regulated under nitrogen starvation conditions, affecting multiple targets such as nitrogenase and

glutamine synthetase, and $sRNA_{162}$ was shown to regulate the switch between carbon and energy sources (Prasse et al., 2017; Jager et al., 2012). Diverse sRNA regulatory mechanisms have also been elucidated in *M. mazei*; $sRNA_{162}$ was reported to bind in *trans* to the ribosome binding site (RBS) of a bicistronic mRNA and to bind in *cis* to the 5' leader region of another mRNA, decreasing the translation of its targets, while $sRNA_{154}$ was shown to bind multiple targets, affecting the stability of those transcripts. A large number of sRNAs have been reported in the halophilic model archaeon, *H. volcanii,* and a few have been assigned potential function, including adaptation to phosphate starvation conditions (Kliemt et al., 2019), nitrogen metabolism (Prasse et al., 2017), and oxidative stress response (Gelsinger & DiRuggiero, 2018b). However, despite the genetic tools available for *H. volcanii* (T Allers et al., 2010; T Allers & Mevarech, 2005; Thorsten Allers et al., 2004), limited sRNA-dependent regulatory mechanisms have been elucidated.

Regulation of the oxidative stress response is particularly relevant for halophilic archaea, which have been found to be extremely resistant to oxidative stress (Baliga et al., 2004; Coker et al., 2007; Gelsinger & DiRuggiero, 2018b; McMillan et al., 2018; Robinson et al., 2011; Sharma et al., 2012), a universal stressor that produces robust phenotypes. The transcription factor RosR was found to be highly expressed under oxidative stress in the haloarchaeon *Halobacterium salinarum*, and to control the expression of over 300 genes (Bidle, 2003; Jevtić et al., 2019). In a previous sRNA-seq screen, we identified hundreds of sRNAs differentially expressed in response to oxidative stress, including both intergenic and antisense sRNAs (Gelsinger & DiRuggiero,

2018b), providing the opportunity to address the mechanistic and functional role of sRNAs in the oxidative stress response of *H. volcanii*.

Here we applied a combination of high throughput and reverse molecular genetic approaches to determine the target, mechanism of action, and functional role of the most up-regulated intergenic sRNA (SHOxi) during oxidative stress in *H. volcanii*. Our work revealed that SHOxi post-transcriptionally regulates an mRNA involved in central metabolism and the production of NADH by facilitating its degradation. As a result of this mRNA destabilization, SHOxi increases the survival of *H. volcanii* by reducing the levels of the dinucleotide NADH, which in turn reduces the damage to macromolecules caused by reactive oxygen species (ROS) inside the cell.

RESULTS

SHOxi is a small non-coding RNA that is responsive to oxidative stress

We previously carried out a sRNA-seq screen in *H. volcanii* under no challenge and oxidative stress conditions and found thousands of differentially expressed sRNAs, which we termed oxidative stress-specific sRNAs (Gelsinger & DiRuggiero, 2018b, **Chapter 2**). In this sRNA-seq screen, we found a novel transcript with no coding capacity (**Fig. 4-S1**) that was enriched 21-fold under oxidative stress conditions (2 mM H_2O_2 exposure for 1 hour; 80% survival) (**Fig. 4-1A**), making this the most up-regulated sRNA in our data set (Gelsinger & DiRuggiero, 2018b). We experimentally confirmed with northern blot analysis that this sRNA transcript was highly expressed under oxidative stress, with only low levels present under no challenge conditions (Fig 1B). The sequence of this sRNA had a high GC content (87%) compared to the average GC content of the *H. volcanii* genome (61.1%) and was located in an intergenic region of the

main chromosome, flanked by two genes on the opposite strand (**Fig. 4-1C**). We identified the transcription start site (TSS) and transcription termination site (TTS) with 5'- and 3'-RACE, resulting in a native sRNA of 234 nt in size. The experimental TSS

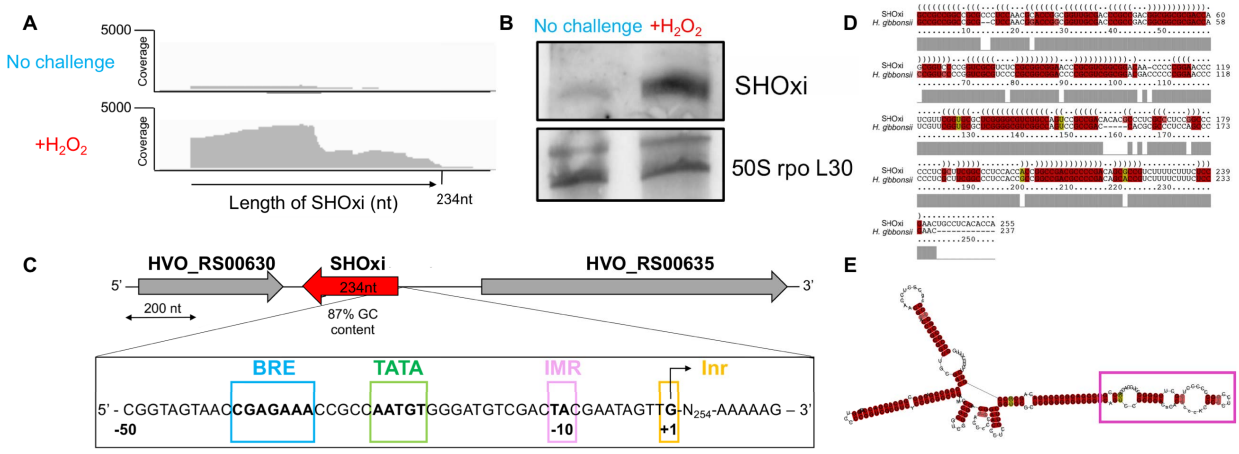

Figure 4-1: Characterization of SHOxi. (A) RNA-seq coverage plots of the assembled SHOxi transcript with data from a previous sRNA-seq screen (Gelsinger & DiRuggiero, 2018b) under no challenge and oxidative stress conditions. (B) *In vivo* verification of SHOxi expression by Northern blot analysis. 50S rpo L30 was used as a loading control. (C) Genomic context of SHOxi. The inlet box is 50nt upstream of the transcription start site of SHOxi and marked are various conserved general transcription motifs that regulate the transcription of SHOxi. (D) Multiple sequence and structural alignment of SHOxi and the sRNA homolog in *H. gibbonsii*. (E) Predicted secondary structure model (minimum free energy) of SHOxi. Putative interaction region highlighted in magenta.

matched the TSS predicted from RNA-seq assembly and allowed for accurate characterization of the basal archaeal transcription factor binding motifs, the B recognition element (BRE), TATA box, initially melted region (IMR), and the initiator element (Inr) upstream (**Fig. 4-1C**).

Using blastn search and the NCBI nt database, we found one highly conserved homolog of SHOxi in *Haloferax gibbonsii,* one of 6 *Haloferax* genomes publicly available. The *H. gibbonsii* sequence was 94% identical at the nucleotide (nt) level, including the upstream regulatory regions, and was located in an intergenic region flanked by two genes with similar function than those in *H. volcanii*. Due to its drastic response to oxidative stress and its high conservation, we named the sRNA **S**mall RNA in **H**aloferax **Oxi**dative stress, or **SHOxi**. We will refer to this sRNA as SHOxi from here on out.

Using SHOxi and its *H. gibbonsii* homologue, we generated a multiple sequence alignment with LocARNA and assessed structural conservation (Fig 1D). We predicted a stable secondary structure (**Fig. 4-1E**) containing high sequence reliability in the first 100 nt and high structural reliability in the last 100 nt, with small drops in reliability in between corresponding to potential loop regions (**Fig. 4-S2**). Although highly structured due to extensive GC base pairing, SHOxi was still predicted to form loops and stem loop regions available for base pairing with mRNA targets (**Fig. 4-1E**).

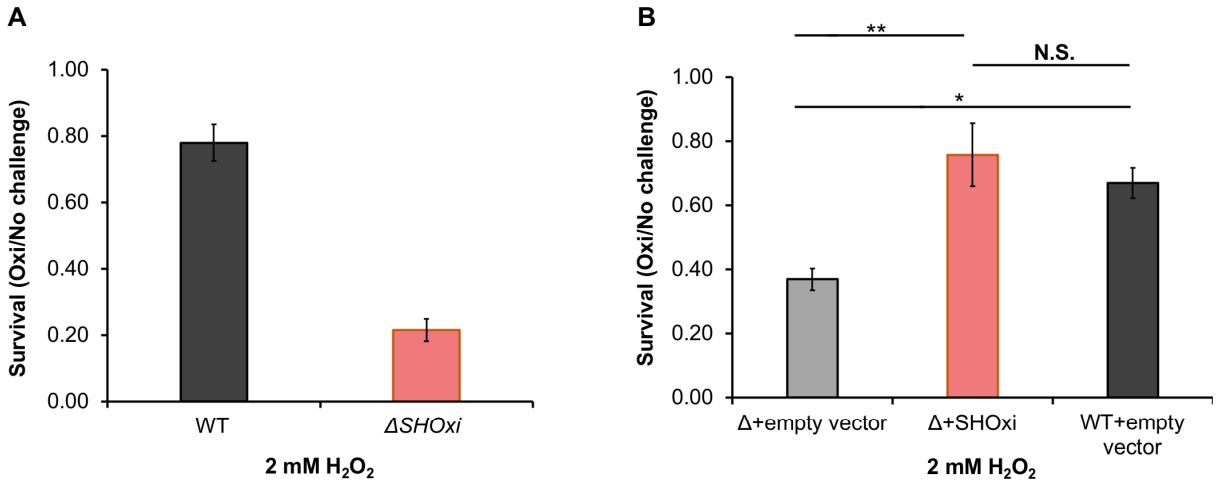

Figure 4-2: Phenotyping of *ΔSHOxi*. (A) Survival of wild type and *ΔSHOxi* under oxidative stress. (B) Rescuing survival by overexpression of SHOxi in a *ΔSHOxi* mutant. The negative control was *ΔSHOxi* (*E*mpty vector) with the empty vector, and the positive control was the wild type with the empty vector (WT+empty vector). SHOxi was overexpressed on the plasmid pTA1300 under a tryptophan inducible promoter in a *ΔSHOxi* background. In both (A) and (B), survival was calculated as the ratio of colony forming units (CFU) between no challenge and oxidative stress conditions (± 2 mM H_2O_2, 1h exposure).

SHOxi increases survival during oxidative stress

To assess whether SHOxi plays a physiological role during oxidative stress in *H. volcanii,* we knocked out the transcriptional loci using a pop-in pop-out method previously established (Thorsten Allers et al., 2004), generating a deletion mutant (*ΔSHOxi*) (**Fig. 4-S3**). We confirmed transcript levels were depleted through northern blot analysis (**Fig. 4-3A**). We then exposed *ΔSHOxi* to 2 mM H_2O_2 for 1 hour to replicate the same oxidative stress conditions as previously described (Gelsinger & DiRuggiero, 2018b) and found *ΔSHOxi* exhibited a drastic decrease in survival (avg. 22% survival) compared to wild type (WT) (avg. 78% survival) (**Fig. 4-2A**).

To test whether SHOxi increased survival during oxidative stress, we constructed an overexpression strain with the SHOxi gene under an inducible tryptophan promoter (pTA1300) (**Fig. 4-S4A**). We confirmed the overexpression of SHOxi with RNA-seq and found a ~32x fold increase of SHOxi levels in both no challenge and oxidative stress conditions, relative to the WT under oxidative stress (**Fig. 4-4SB**). WT *H. volcanii* transformed with an empty vector was used as a positive control and yielded ~ 67%

survival during oxidative stress (+2 mM H_2O_2, 1h). As a negative control, *ΔSHOxi* transformed with an empty vector yielded low survival (avg. 37% survival) during oxidative stress. However, ectopic expression of a SHOxi plasmid in a *ΔSHOxi* background resulted in rescued survival levels (avg. 76% survival) under oxidative stress conditions, comparable to WT (**Fig. 4-2C**).

SHOxi changes the expression of many putative target mRNAs

To identify SHOxi mRNA targets, we sequenced the transcriptome of *ΔSHOxi* and WT *H. volcanii* under oxidative stress conditions, and WT *H. volcanii* under no-challenge conditions. We found 215 mRNAs with significant log2 fold-changes (≥2) and with a false discovery rate less than 5% between *ΔSHOxi* and WT during oxidative stress (**Fig. 4-3B**). We further restricted these putative targets to only include mRNAs with opposite fold change patterns (≥2) between WT oxidative stress and WT no-challenge conditions, to increase the stringency of our analysis and the potential of selecting mRNA targets affected by SHOxi and no other factors (i.e. oxidative stress). Using these stringent criteria, 46 putative targets of SHOxi were identified (**Fig. 4-3B, red dots**). A gene ontology analysis (DAVID) found putative targets up-regulated in the absence of SHOxi were significantly ($p<0.05$) enriched for transcriptional regulators, while down-regulated targets were enriched for sugar metabolism.

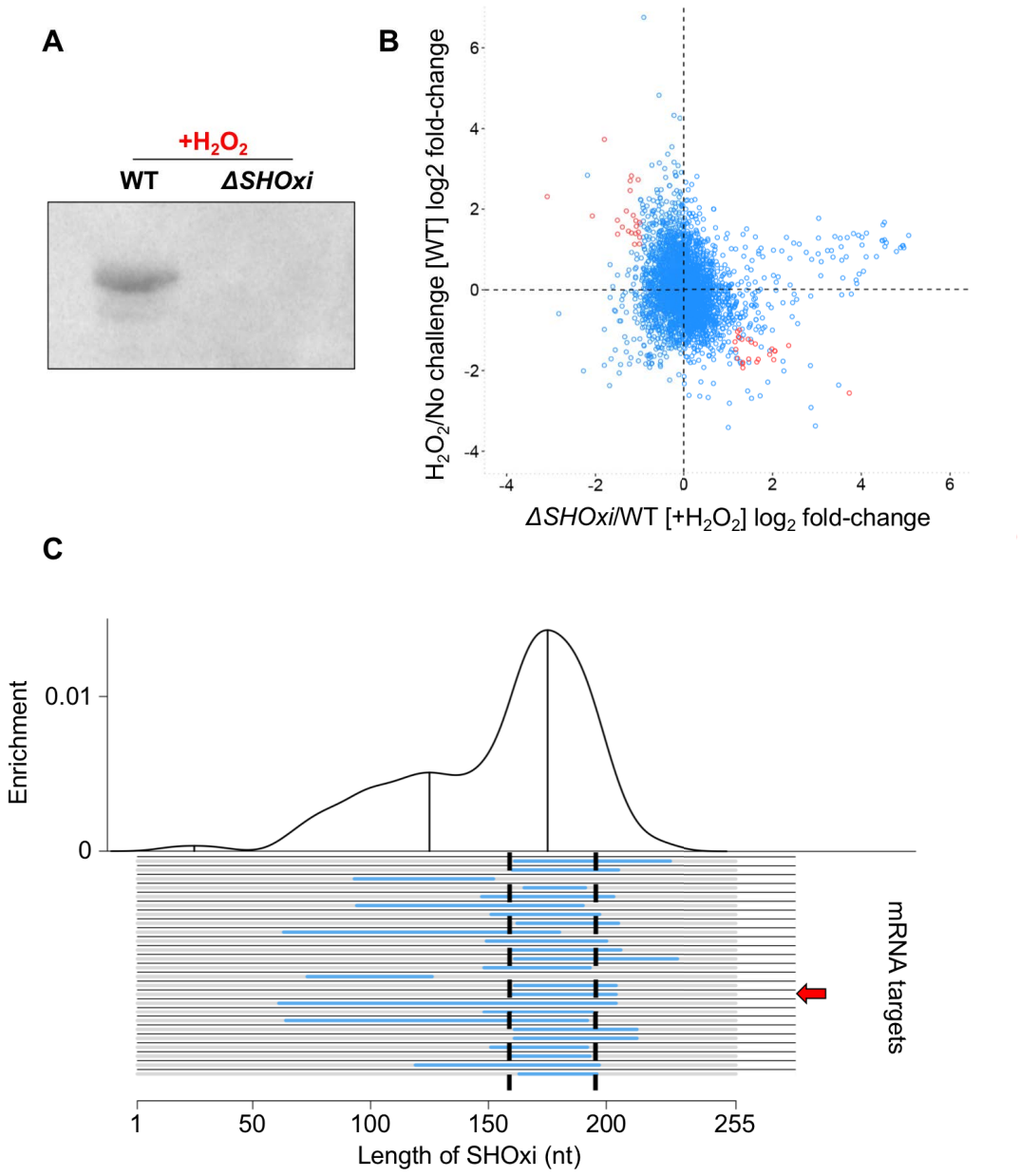

Figure 4-3: Identification of potential targets of SHOxi. (A) Confirmation of depleted SHOxi transcript levels via Northern blot analysis. The band corresponds to SHOxi transcript. (B) Scatterplot of differentially expressed genes in the absence of SHOxi (fold-change between *ΔSHOxi* +H_2O_2 and WT +H_2O_2) compared to the wild type under oxidative stress (fold-change between +H_2O_2 and no challenge). (C) Predicted hybridization plot of the top 25 most probable interactions in the entire transcriptome of *H. volcanii*.

An *in-silico* approach was also used to find (1) interacting partners based on hybridization (sRNA-mRNA interactions) and (2) whether there was a region within SHOxi that seems most probable for interaction (i.e. lowest free energy change). We used IntaRNA to calculate hybridization energies between SHOxi and all the transcripts in the NCBI *H. volcanii* annotation. This analysis yielded a 20 nt conserved region in SHOxi that was putatively assigned as the interaction site for the 25 most reliably predicted targets ($p<0.01$) (**Fig. 4-3C**). This putative interaction site corresponded to a multi stem loop region in the modeled secondary structure of SHOxi (**Fig. 4-1E, magenta**).

By intersecting our *in-silico* and experimental approaches to identify SHOxi targets, we found one transcript mRNA that was significantly up-regulated (FDR = 5.48E-12) in *ΔSHOxi* and was predicted to have strong RNA-RNA interaction with SHOxi (**Fig. 4-4A**). This transcript encoded a NAD-dependent malic enzyme mRNA, which is known to convert NAD+ to NADH. SHOxi was predicted to interact with a 40 nt region, ~300 nt downstream of the TSS of this mRNA, with a significantly strong hybridization energy (-31 kcal/mol, $p = 0.00128$) (**Fig. 4-3C, red arrow**). The region of interaction corresponded to the predicted stem loop interaction site and a putative "seed" region (a segment of contiguous base-pairing) of SHOxi at 166 to 174 nt within the SHOxi stem loop (**Fig. 4-4A**).

The stem loop region of SHOxi interacts directly with malic enzyme mRNA to regulate its expression

sRNAs in all three domains of life mostly interact with their targets via RNA-RNA base pairing, which in turn can affect the transcript levels of the target mRNA (Carthew & Sontheimer, 2009; Gottesman & Storz, 2011; Storz et al., 2011). Using our

RNA-seq dataset, we found that malic enzyme mRNA levels decreased 2-fold in the WT under oxidative stress and increased a little more than 2-fold in *ΔSHOxi* under oxidative stress, when compared to no challenge conditions (**Fig. 4-4B**).

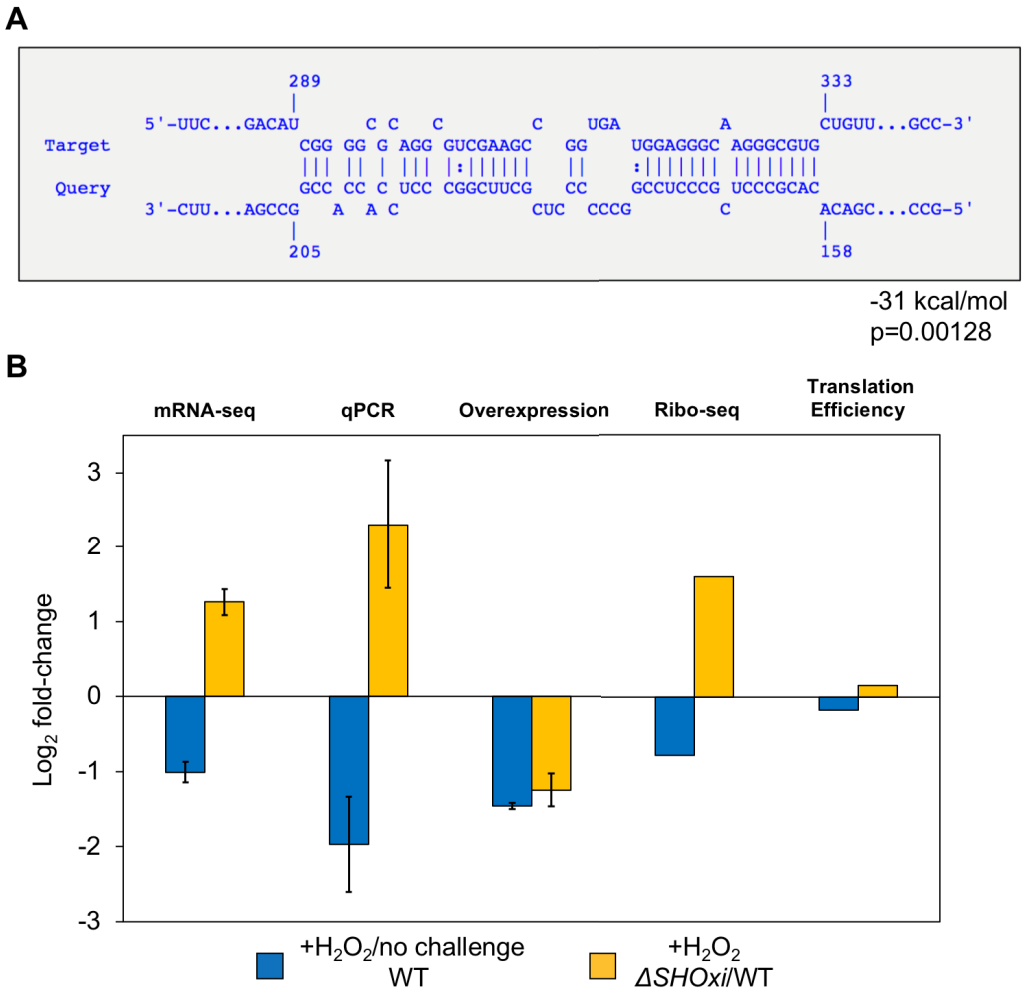

Figure 4-4: Regulation of malic enzyme mRNA by SHOxi. (A) Predicted interaction between malic enzyme mRNA (target) and SHOxi (query). Wobble base bonds indicated by two dots between hybridizing nucleotides. (B) Bar plot of expression profiles of the mRNA level (mRNA-seq, qPCR, overexpression), and translation (Ribo-seq, TE).

This strongly suggested that malic enzyme mRNA might be a direct target of SHOxi. The RNA-seq results were validated using quantitative (q)RT-PCR in WT, *ΔSHOxi*, and in our constructs overexpressing SHOxi in a WT or *ΔSHOxi* background, under oxidative stress (+2 mM H_2O_2, 1h) (**Fig. 4-4B**). While the alteration of malic enzyme transcript levels may play a role in the decreased survival of *H. volcanii*, transcript levels may also affect translation. We have recently developed ribosome profiling in *H. volcanii* (Gelsinger et al., 2020), a global measure of translation in a cell, and carried out ribosome profiling on WT and *ΔSHOxi* under no-challenge and oxidative stress conditions. We found that malic enzyme mRNA translation levels correlated with transcription levels (**Fig. 4-4B**). Additionally, translation efficiency measurements reveal no significant difference between WT and *ΔSHOxi* under oxidative stress (**Fig. 4-4B**). This indicates SHOxi's regulatory effect on malic enzyme mRNA is most likely upstream of translation and is post-transcriptionally potentially mediated by direct RNA-RNA interactions.

To establish the SHOxi sRNA and malic enzyme mRNA interaction sites, we used site-directed mutagenesis to construct 3 SHOxi mutants with various mutations in the predicted "seed" binding region (**Fig. 4-5A**). Computational predictions revealed that 2 out of 3 SHOxi mutants substantially disrupted the binding between the SHOxi and malic enzyme RNA seed region ($\Delta\Delta G < -2.0$ kcal/mol) (**Fig. 4-5B**). The SHOxi mutant constructs were experimentally tested by overexpression in a *ΔSHOxi* background under no challenge conditions. Instead of using oxidative stress to induce SHOxi expression, the SHOxi mutant sRNAs were placed under an inducible promoter so we could simulate the SHOxi expression by the addition of tryptophan to the cultures (+2 mM tryptophan for 1h). *In vivo* malic enzyme mRNA levels were measured using qRT-PCR and fold

changes were calculated relative to the overexpression of the non-mutated SHOxi in a *ΔSHOxi* background under no challenge conditions.

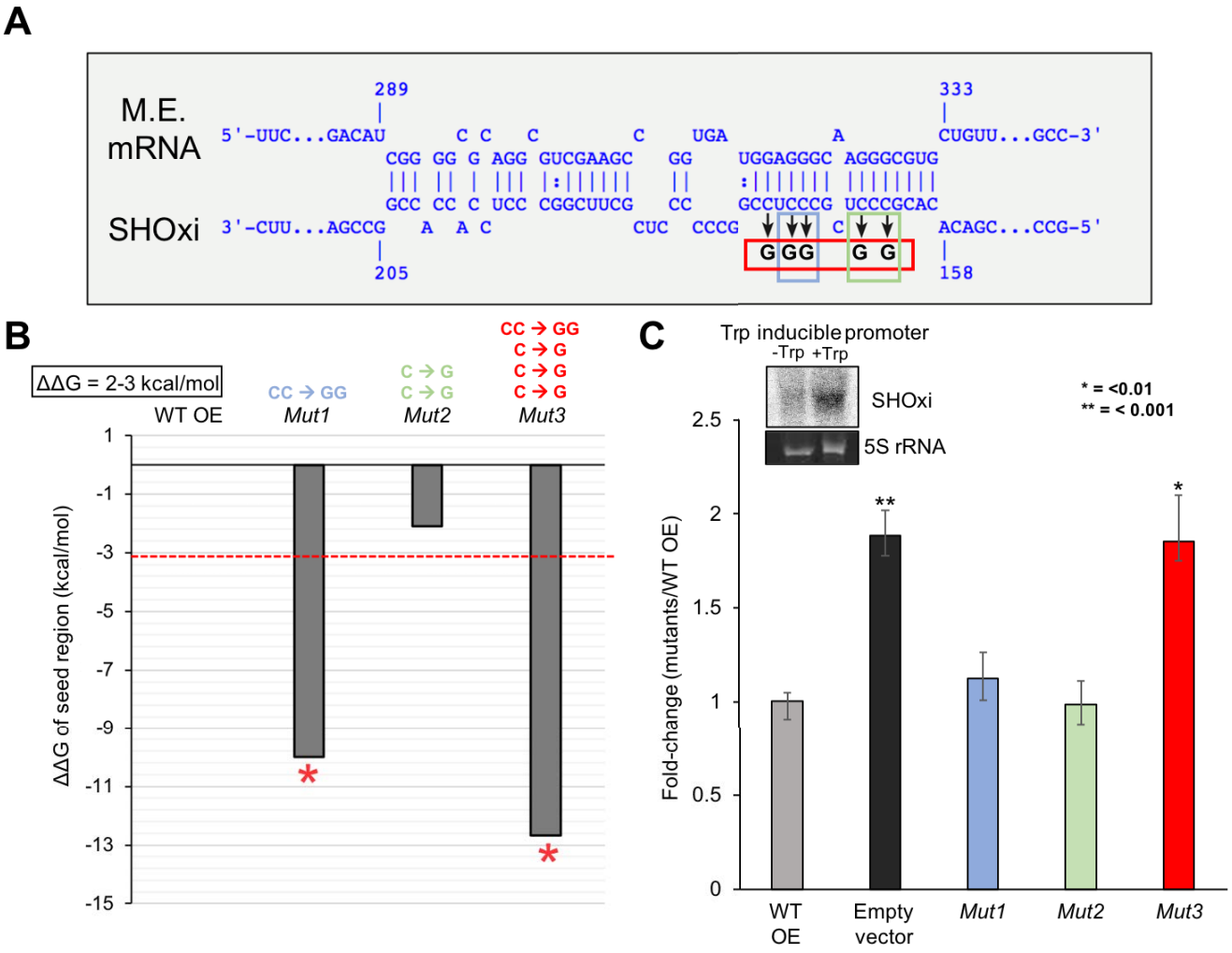

Figure 4-5: *in vivo* validation of direct interaction between SHOxi-malic enzyme mRNA. (A) Schematic of mutations in the "seed" binding region between SHOxi and malic enzyme mRNA. (B) Free energy changes in the binding of each SHOxi mutation variant with malic enzyme mRNA. Red dashed line denotes a free energy change of -2 kcal/mol. (C) qRT-PCR on malic enzyme mRNA levels, calculated as the fold change relative to overexpression of the WT SHOxi.

The negative control, *ΔSHOxi* harboring the empty vector, exhibited higher malic enzyme mRNA levels compared to the positive control, *ΔSHOxi* harboring WT SHOxi on a plasmid. We found that point mutations or dinucleotide mutations in the "seed" binding region of SHOxi did not alter expression of malic enzyme (**Fig. 4-5C, Mut1 and Mut2**), indicating a non-disruptive effect in binding. In contrast, a combination of 5 nucleotide mutations in the "seed" binding region of SHOxi resulted in the increase of malic enzyme mRNA levels, comparable to that of the negative control, indicating a disruptive effect in binding between the sRNA and mRNA (**Fig. 4-5C, Mut3**). This result confirms our prediction for the binding interaction between malic enzyme mRNA and the stem loop region of SHOxi and provides a path for the regulation of malic enzyme mRNA.

Malic enzyme mRNA stability is post-transcriptionally regulated by SHOxi

To further verify that SHOxi directly affected malic enzyme transcript levels, we measured malic enzyme transcript stability *in vivo*. Following SHOxi transcription induction by a 30 min exposure to H_2O_2, we added Actinomycin D (actD) to WT and *ΔSHOxi* cultures to inhibit transcription. We then extracted total RNA at time intervals of 0, 15, 30, and 60 min after actD addition and used qPCR to measure malic enzyme mRNA levels (**Fig. 4-6A**). A house keeping gene with no altered expression level in WT and *ΔSHOxi*, a ribosomal protein (rpl30), was used as a control. We found that malic enzyme mRNA transcript levels were higher over the time course in *ΔSHOxi* compared to WT under oxidative stress, indicating the mRNA was more stable in *ΔSHOxi* (**Fig. 4-6B**). In contrast, the surface glycoprotein mRNA did not exhibit any significant difference in transcript levels between *ΔSHOxi* and WT under oxidative stress (**Fig. 4-6C**).

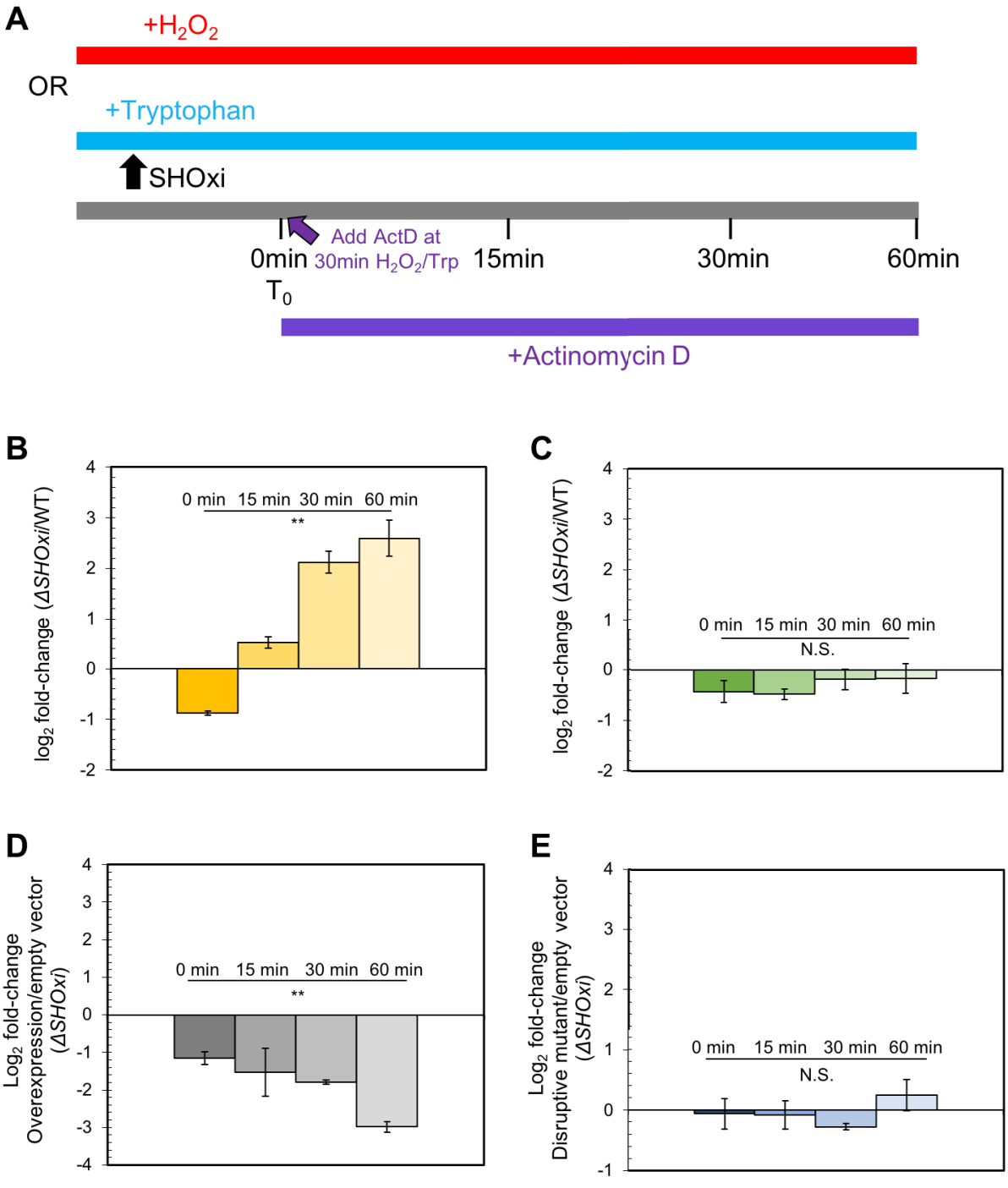
A
+H2O2
OR
+Tryptophan
SHOxi
Add ActD at
30min H2O2/Trp
0min
T0
15min
30min
60min
+Actinomycin D
B
log2 fold-change (ΔSHOxi/WT)
0 min 15 min 30 min 60 min
**
C
log2 fold-change (ΔSHOxi/WT)
0 min 15 min 30 min 60 min
N.S.
D
Log2 fold-change
Overexpression/empty vector
(ΔSHOxi)
0 min 15 min 30 min 60 min
**
E
Log2 fold-change
Disruptive mutant/empty vector
(ΔSHOxi)
0 min 15 min 30 min 60 min
N.S.

Figure 4-6: Destabilization of malic enzyme mRNA. (A) Schematic of the experimental approach to measure the half-life of malic enzyme mRNA in response to SHOxi. SHOxi was induced by treating the cells (WT and *ΔSHOxi*) with 2 mM H_2O_2 or SHOxi and the disruptive SHOxi mutant from **Fig. 4-5C** (in *ΔSHOxi*) were overexpressed using an inducible promoter in no challenge conditions by addition of 2 mM tryptophan. After 30 min, transcription was shut off by the addition of 100 ug/ml actinomycin D and the level of malic enzyme mRNA was analyzed by qRT-PCR at several time points. (B) qRT-PCR of malic enzyme mRNA levels over time after shutting off transcription. Log_2 fold changes are calculated between the WT and *ΔSHOxi* with 2 mM H_2O_2. (C) qRT-PCR of the house keeping gene surface protein mRNA levels in the same conditions as in (B). (D) qRT-PCR of malic enzyme mRNA levels after shutting off transcription in response to the overexpression of SHOxi with 2 mM tryptophan under no challenge conditions. Log_2 fold changes are calculated between the SHOxi overexpression construct and the empty vector in a *ΔSHOxi* background. (E) qRT-PCR of malic enzyme mRNA levels in the same conditions as in (D) but with the overexpression of the disruptive SHOxi mutant.

Next, we simulated the expression of SHOxi due to oxidative stress by the addition of tryptophan to the cultures (*ΔSHOxi* background) in no challenge conditions (+2 mM tryptophan for 1h), inhibited transcription with actD, and measured malic enzyme mRNA levels under the same time course as (**Fig. 4-6 B-C**). We found that malic enzyme mRNA transcript levels were lower over time in the overexpression of SHOxi compared to the empty vector control (both in a *ΔSHOxi* background), despite no oxidative stress challenge (**Fig. 4-6 D**). In contrast, in the overexpression of the disruptive

SHOxi mutant from **Fig. 4-5C** did not exhibit any significant difference in malic enzyme mRNA transcript levels compared to the empty vector (**Fig. 4-6 E**). These findings demonstrated that malic enzyme mRNA was destabilized by SHOxi during oxidative stress (**Fig. 4-6**).

NADH levels are regulated by SHOxi during oxidative stress

NAD-dependent malic enzyme converts NAD+ to NADH. To test whether malic enzyme mRNA destabilization by SHOxi had a direct effect on NADH levels, we measured all nicotinamide adenine dinucleotides in the cell (NAD+, NADH, NADP+, NADPH), for both WT and *ΔSHOxi* during no challenge and oxidative stress conditions, using a luciferase-based assay. We then calculated ratios between NAD^+:NADH and $NADP^+$:NADPH levels to normalize the results between different conditions and strains. We found the NAD^+:NADH ratio increased under oxidative stress (when SHOxi is up-regulated) relative to no challenge and *ΔSHOxi* conditions (**Fig. 4-7A**). In contrast, the ratio of $NADP^+$:NADPH was not altered for any of the experimental conditions (**Fig. 4-7A**).

Cellular respiration driven by the electron transport chain, a process that ultimately converts NADH to ATP, is the main source of intracellular reactive oxygen species (ROS) within cells. We hypothesized that the decrease of NADH levels via SHOxi could alter redox homeostasis in *H. volcanii*. Using western blotting, we measured the amount of protein carbonyl groups as an indicator of ROS-mediated damage to protein side chains. To validate our assay, we first measured the amount of WT *H. volcanii* protein carbonyl groups under oxidative stress and in no-challenge conditions (**Fig. 4-7B**). When measuring *ΔSHOxi H. volcanii* protein carbonylation under

oxidative stress, we found an increase of ~1.64X compared to WT, indicating increased oxidative damage likely as the result of more ROS production (**Fig. 4-7C**).

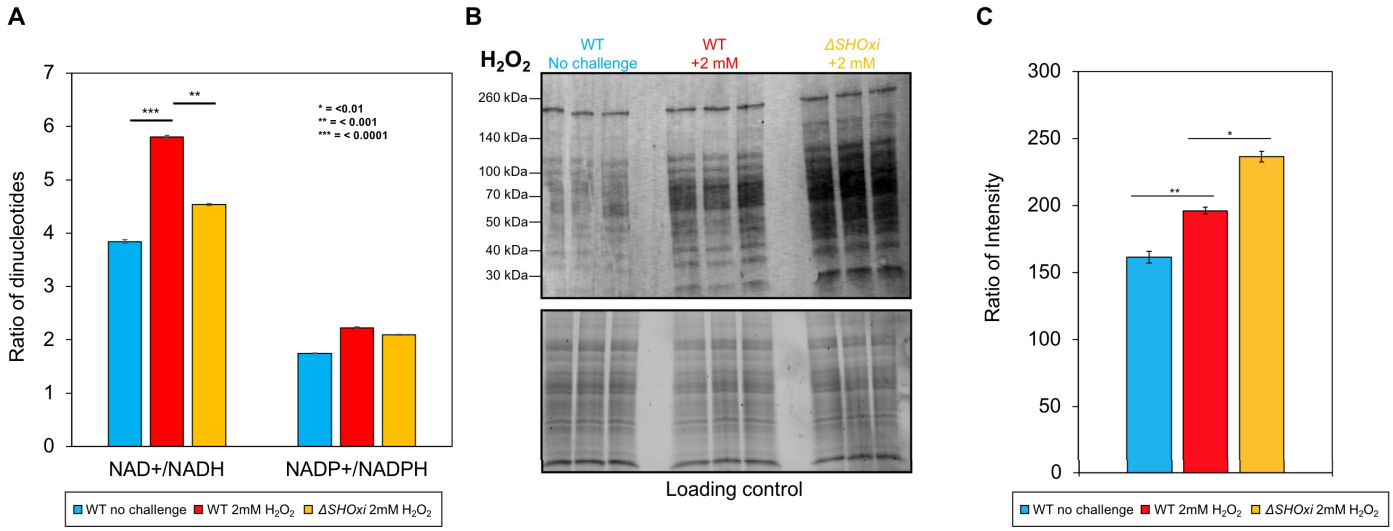

Figure 4-7: Functional consequences of malic enzyme mRNA destabilization during oxidative stress. (A) Measurements of all dinucleotides using a luciferase based assay. Ratios of NAD+/NADH and NADP+/NADPH were calculated to assess the relative abundance. (B) Oxyblot western blot on carbonyl groups found on proteins, a proxy for oxidative damage. Samples were the wild type under no challenge conditions (0 mM H_2O2), the wild type exposed to 2 mM H_2O_2 (80% survival), and *ΔSHOxi* exposed to 2 mM H_2O2. A loading control gel is provided blow the plot. (C) Quantification of (B). Legends correspond.

DISCUSSION

By unraveling the mechanism and function of the most up-regulated intergenic sRNA, SHOxi, during H_2O_2-induced oxidative stress in *H. volcanii,* we found that this sRNA is a major regulator of the oxidative stress response. SHOxi is highly conserved in another *Haloferax* species, it is predicted to form a highly structured conformation, and was enriched 21-fold under oxidative stress. Deletion of SHOxi (*ΔSHOxi*) caused a drastic

decrease in survival of *H. volcanii* under oxidative stress compared to the WT, underlying a key role in the stress response.

From our mRNA-seq data, we identified several putative targets that were differential expressed in presence (WT) and absence (*ΔSHOxi*) of SHOxi. These putative targets included several transcription factors with unknown functions, suggesting that SHOxi may be a master regulator with large downstream effects in the gene regulatory network. Interestingly, all of these transcription factor mRNAs had increased transcript levels in the presence of SHOxi, suggesting they might be stabilized by the sRNA, instead of degraded like the malic enzyme mRNA. This would not be unusual since dual functioning sRNAs have been observed in other Archaea (Jager et al., 2012; Prasse et al., 2017). In contrast, several other putative targets were down-regulated in the presence of SHOxi, such as a sugar ABC transporter operon.

In contrast to Archaea, several sRNAs have been implicated in the regulation of oxidative stress in Bacteria (Altuvia et al., 1997; Barshishat et al., 2018; Chen et al., 2019; Lalaouna et al., 2019). The bacterial sRNA OxyS is the best studied and its regulatory mechanism has recently been fully characterized (Barshishat et al., 2018). OxyS protects *Escherichia coli* cells from DNA damage by decreasing the translation of the essential transcription termination factor NusG. This leads to an increase of the virulence factor kilR, which, in turn, interferes with the cell division protein FtsZ and ultimately inhibits cell division. The arrest in cell growth provides more time for DNA damage repair, hence better survival under oxidative stress (Barshishat et al., 2018). To date, no published work has implicated an sRNA in the regulation of oxidative stress in Eukarya. Here we report, for the first time in the Archaea, the functional role of a target for an oxidative stress-

responsive sRNA. Malic enzyme mRNA, a highly expressed transcript (11th percentile of expressed mRNAs), which encodes for a protein involved in the conversion of NADH to ATP through central metabolism, was significantly down-regulated in the presence of SHOxi. We demonstrated that the malic enzyme mRNA had a strong RNA-RNA interaction through extensive base pairing of the stem-loop region in SHOxi's secondary structure. By measuring steady state RNA levels via actinomycin D treatment, we showed that malic enzyme mRNA had decreased stability over time only when SHOxi was present in the cell. Thus, our findings indicate that the mechanism by which SHOxi might regulate malic enzyme during oxidative stress is by destabilizing its mRNA through direct RNA-RNA binding.

Destabilization of mRNAs through RNA-RNA interactions and recruitment of a RNase has been well documented in Bacteria (Baek et al., 2019; Bandyra et al., 2012). We speculate here that the destabilization of malic enzyme mRNA by SHOxi is carried out by a currently unknown RNase. RNases and RNA degradation pathways are still not well resolved in Archaea, although some recent studies have brought insights onto potential biochemical mechanisms for novel RNases (Clouet-d'Orval et al., 2018; Levy et al., 2011; Randau, 2012; Wurtmann et al., 2014; Yue et al., 2019). Moreover, due to the internal interacting site of the malic enzyme mRNA, we speculate that an endonuclease would be the most likely RNase. Intriguingly, a RidA endonuclease was found to play a functional role in stress response in *H. salinarum* (Wurtmann et al., 2014). The corresponding RidA homolog in *H. volcanii* is indeed the most up-regulated gene during oxidative stress, which may indicate a crucial role in RNA processing during oxidative stress (Gelsinger & DiRuggiero, 2018b).

Lastly, we found that the functional role of SHOxi-mediated post-transcriptional regulation of malic enzyme mRNA was to decrease NADH levels, which, in turn reduced the amount of damage to the cell's macromolecules, including its proteins, during oxidative stress. NAD^+/NADH and $NADP^+$/NADPH redox couples are essential for maintaining cellular redox homeostasis in the cell (Xiao et al., 2018). NAD^+ is involved in catabolic reactions including glycolysis and the tricarboxylic acid (TCA) cycle, while $NADP^+$ serves as an electron donor for reductive biosynthesis. NADPH also acts as a reducing agent to regenerate antioxidant systems such as thioredoxins and glutathione. Under no challenge conditions, NADH, generated primarily in the TCA cycle, is oxidized to NAD^+ at the electron transport chain (ETC) by a NADH dehydrogenase. Electrons from this oxidation are shuttled along the ETC to ultimately reduce oxygen to water. This process is coupled with the generation of a proton gradient and ATP synthesis (Mountfort, 1978). Under H_2O_2-mediated oxidative stress, the level of ROS in the cell dramatically increases as the result of the oxidation of methionine and cysteine residues and the attacks of iron-sulfur cluster proteins in the presence of transition metal ions (Berlett & Stadtman, 1997; Bush, 2000; Giles et al., 2003; Stadtman, 1993). The subsequent Fenton-like reactions result in enzyme inactivation, and the proliferation of ROS via the release of Fe^{2+}, interfering directly with essential metabolic pathways in the cell (Imlay et al., 2003, 2006, 2008).

We propose here that in *H. volcanii* during no challenge conditions, SHOxi is not induced and, thus, limited interaction occurs between the sRNA and malic enzyme mRNA. Optimal amounts of NADH is produced to convert ADP into ATP and corresponding amounts of NADPH is produced to maintain redox homeostasis (**Fig. 4-8A**). Under oxidative stress, by limiting the production of NADH, via the destabilization of malic

enzyme mRNA, SHOxi prevents additional ROS production, thus, limiting additional oxidative damage to the cell's macromolecule (**Fig. 4-8B**). In a SHOxi knockout mutant (*ΔSHOxi*), NADH is still produced, resulting in overload and leakage of the ETC, increase production of ROS, and a severe decrease in survival (**Fig. 4-8C**).

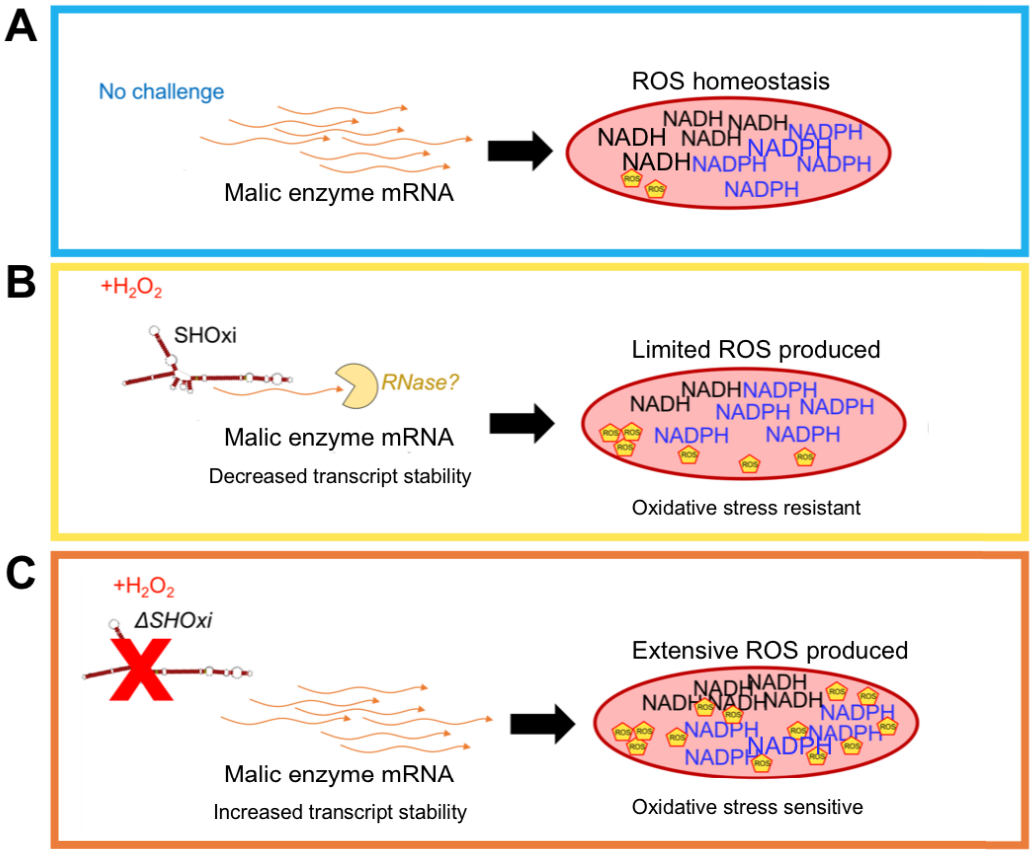

Figure 4-8: A model for the posttranscriptional regulation of malic enzyme mRNA by SHOxi. An illustration depicting the different models of malic enzyme mRNA regulation during oxidative stress in the presence or absence of SHOxi. (A) Wild type no challenge conditions and the low expression profile of SHOxi. ROS homeostasis is mentioned. (B) Wild type under oxidative stress conditions (+2 mM H_2O_2, 80% survival). SHOxi is up-regulated and ROS production is limited. (C) A knockout of SHOxi (*ΔSHOxi*) under the same conditions as (B). No expression of SHOxi and extensive ROS-mediated damage is produced.

This sRNA-mediated regulation of redox homeostasis via the modulation of NAD^+/NADH ratio is most likely part of the global cellular response to oxidative damage that includes the upregulation of enzymatic detoxification systems (superoxide dismutases, catalases, peroxidases) and antioxidants (glutathione) (Forman et al., 2009), and the down-regulation of metabolism, as it has previously been reported in haloarchaea (Baliga et al., 2004; Bonneau et al., 2007; Facciotti et al., 2010; Gelsinger & DiRuggiero, 2018b; Sharma et al., 2012). Alternatively, an increase in cellular NAD^+, as the result of SHOxi activation under oxidative stress, could provide additional template for the enzymatic conversion of NAD^+ to $NADP^+$ by an NAD kinase, generating additional NADPH, with strong antioxidant properties (Shi et al., 2009; Yang & Sauve, 2016). However, we did not find evidence for a change in the ratio of $NADP^+$:NADPH under oxidative stress conditions in *H. volcanii*.

Haloarchaea in their natural environment are exposed to intense solar radiation and desiccation, all of which generate high levels of oxidative stress (Oren, 2014), and thus need fast responses to extreme environmental stresses. sRNAs have the potential to carry out such a role by fine-tuning the regulation of genes. The oxidative stress response in *H. volcanii* and *H. salinarum* was shown to impact a wide array of cellular processes, engaging at least 50% of all the genes (Gelsinger & DiRuggiero, 2018b; McMillan et al., 2018; Whitehead et al., 2006). These changes were characterized by the up-regulation of DNA repair (RpaA/B genes), ROS scavenging enzymes (e.g. catalase), protein turnover, and iron sulfur assembly proteins, and the down-regulation of metabolism (Baliga et al., 2004; Bonneau et al., 2007; Facciotti et al., 2010; Gelsinger & DiRuggiero, 2018b; Sharma et al., 2012). Moreover, both transcription factors and sRNAs have been shown to be

responsive to oxidative stress in haloarchaea providing further evidence that the response to this stress is highly regulated (Gelsinger & DiRuggiero, 2018b; Sharma et al., 2012). This underlines the importance of SHOxi, a functional non-coding RNA, in the resistance of an extremophile to oxidative stress. Ultimately, this study establishes the regulatory effects of an sRNA on mRNAs during the oxidative stress response in the Archaea. Elucidating the post-transcriptional regulation of malic enzyme mRNA by the sRNA SHOxi provides evidence that a sRNA is crucial for the survival of an archaeon to stress and that fine-tuning of metabolism is a core strategy to mitigate damage from oxidative stress. Future work will determine whether an RNase is involved in the destabilization mechanism and if RNA binding protein help facilitate interactions between SHOxi and mRNA targets.

MATERIAL AND METHODS

Culture growth conditions. *H. volcanii* auxotrophic strain H53 (*Δpyre2, ΔtrpA*) and H98 (*Δpyre2, ΔthyH*) were used for all experiments. Culturing in liquid and solid media was done in rich medium (Hv-YPC) or selection medium (Hv-Cab), at 42°C and with shaking at 220 rpm (Amerix Gyromax 737). Uracil, tryptophan, thymidine, and hypoxanthine were added to a final concentration of 50 µg/mL, each.

Knock out generation. Deletion mutants of SHOxi (ΔSHOxi) were constructed using the pop-in pop-out method described previously (Thorsten Allers et al., 2004), independently in H53 and H98 strain backgrounds. 500 base pairs upstream and downstream of SHOxi including small overhangs (30 bp) were PCR amplified, stitched together, and then cloned into the integration vector pTA131 to create a knockout vector. *H. volcanii* strain

H53 and H98, respectively, were transformed with the plasmid to yield pop-in clones by uracil autotrophy. To generate the pop-out strains, these cells were plated on medium containing 5-fluoro-orotic acid (5-FOA). Deletions were verified at the DNA level by PCR and at the RNA level by northern blot and RNA-seq.

Oxidative stress exposure. *H. volcanii* liquid cultures were exposed to the oxidative stress agent H_2O_2 as previously described (Gelsinger & DiRuggiero, 2018b). Cultures were grown in 160 mL of Hv-YPC or Hv-Cab under optimal conditions to an OD of 0.4 (mid exponential phase). 2 mM H_2O_2 was directly added to the cultures followed by an hour incubation at 42 °C with shaking at 220 rpm. Cultures were then rapidly cooled down, centrifuged at 5,000 x g for 5 minutes and the pellets resuspended in 18% sea water. The cell suspensions were then transferred to a 1 mL tube and centrifuged at 6,000 x g for 3 minutes, the pellets were flash frozen and stored at -80 °C until ready for RNA extraction.

RACE analysis. The 5' and 3' ends of SHOxi were determined using the Takara SMARTer Rapid Amplification of cDNA Ends (RACE) kit with slight modifications on total RNA extracts from oxidative stress treated cells previously described. For 5' RACE the protocol for cDNA generated by random primers was used, followed by the standard protocol with custom internal reverse primer complementary to SHOxi. For 3' RACE, total RNA was treated with polyA polymerase (NEB) for 1h at 37 °C to add polyA-tails to RNAs. Afterwards, the standard 3' RACE protocol was followed.

Overexpression experiments. A variant of the overexpression plasmid pta1228 (T Allers et al., 2010) was created in order to prevent the introduction of an ATG at the beginning of SHOxi. Using the standard protocol of the Q5 Site-directed Mutagenesis kit (NEB) the region of pta1228 spanning restriction sites *NdeI* and *BamHI* were replaced with *KasI*

yielding the new plasmid pta1300. The full length of SHOxi was PCR amplified with overhangs and sticky end ligated at the *KasI* restriction site of pta1300. *H. volcanii* was transformed using the pop-in method and uracil autotrophy to generate both H53 and H98 wild-type and *ΔSHOxi* overexpression clones. Overexpression was induced at OD 0.4 by addition of 2mM tryptophan and incubation at 42 °C with shaking at 220 rpm for 1 hour. Cells were harvested as described above and used for mRNA-seq or qPCR.

Oxidative stress survival curves. Assessment of survival in *H. volcanii* wild type and ΔSHOxi under oxidative stress conditions was done using microdilution plating as described. Counts were averaged and standard deviation calculated between replicates. Survival was calculated as the number of viable cells following H_2O_2 treatment divided by the number of viable untreated cells and graphed with standard error bars.

RNA extraction. Total RNA was extracted using the Zymo Quick-RNA Miniprep kit with the following modifications: after addition of RNA lysis buffer to the frozen cell pellets, cells were processed with a 23 G needle and syringe to insure complete cell lysis. *H. volcanii* liquid culture is slimy and viscous thus to increase cellular lysis a 23 G needle and syringe was used to break down the cell pellet. Total RNA was then extracted following the standard kit protocol.

Messenger RNA-sequencing library preparation (mRNA-seq). Total RNA was DNase I (NEB) treated (37 °C for 2 hours) as previously described. Total RNA was then rRNA-depleted using the Ribo-zero Bacteria kit (Illumina). Strand-specific libraries were prepared using the SMART-seq Ultralow RNA input kit (Takara), insert sizes checked with the Bioanalyzer RNA pico kit (Agilent), and either paired-end sequencing (2 x 150 bp) or

single-end sequenced (100 bp) was carried out on the Illumina HiSeq 2500 platform at the Johns Hopkins University Genetic Resources Core Facility (GRCF).

mRNA-seq differential expression analysis. We used a read count-based differential expression analysis to identify putative targets of SHOxi that are differentially expressed during oxidative stress and when SHOxi is knocked out. The program featureCounts was used to rapidly count reads that map to the NCBI *H. volcanii* annotation. featureCounts was run with strand-specific options on, paired-end mode on or off, multi-mapping off. The read counts were then used in the R differential expression software package DESeq2. Briefly, read counts were converted into a data matrix and normalized by sequencing depth and geometric mean. Differential expression was calculated by finding the difference in read counts between the SHOxi knockout oxidative stress state to the normalized read counts from the wild-type oxidative stress normalized read counts. The differentially expressed mRNAs were filtered based on the statistical parameter of False Discovery Rate (FDR) under 5%. In addition, only mRNAs with converse differential expression levels (FDR <%) in our previous wild type no challenge/oxi stress differential expression comparison (Gelsinger & DiRuggiero, 2018b) were labeled as specific putative targets of SHOxi.

Northern Blot analysis. 20 μg of total RNA and P^{32} ATP end-labeled Century+ RNA markers were loaded onto 5% denaturing urea polyacrylamide gels (SequaGel, National Diagnostics) and run at 30 watts for 1.5 hours to ensure well-spaced gel migration from 50 to 1,000 nucleotides (nt). Gels were transferred onto Ultra-hyb Nylon membranes and hybridized with 2 types of probes. For SHOxi, we probed with [γ-P^{32}] dATP randomly primed amplicons generated with custom primers. Probe primers were at a minimum 10 nt

inwards from the predicted genomic coordinates (start and stop) to ensure accurate transcript detection. Hybridizations were done at 65°C. For mRNAs (malice enzyme, etc.) we used [α-P^{32}] dATP end-labeled oligo probes (20-23 nt). Hybridizations were done at 42°C. The rpl30 protein (HVO_RS16975) transcript was used as a loading control for differential expression calculation because it was not differentially expressed under oxidative stress in our previous RNA-seq dataset. Differential expression was calculated using ImageJ.

***In silico RNA interactions*.** The program IntaRNA (Busch, Richter, & Backofen, 2008) was used to computationally predict possible interactions with SHOxi and all RNAs in the NCBI *H. volcanii* gene annotation. Options used were no-seed, 42 °C temperature, no START. Top candidates were the top 100 hits ranked by lowest p-value.

RNA half-life measurement. wild type or *ΔSHOxi* cells at OD 0.4 were grown for 30 min with H2O2 to induce endogenous expression of SHOxi and subsequently treated with actinomycin D to inhibit transcription. Samples were harvested at 0,15,30, and 60 minutes post-actD which were extracted for RNA malic enzyme mRNA levels were measured with qRT-PCR 0, 15, 30, and 60 minutes post-actD, in *ΔSHOxi* and wild type, under oxidative stress.

Binding site mutagenesis experiments. Mutations within the stem-loop binding site of SHOxi were constructed using the Q5 Site-directed Mutagenesis kit (NEB) standard protocol on the previously described SHOxi overexpression construct (in pta1300). The forward primer (CCGACACACGGCGTCGCGGTGCGGCCCCCCT) and reverse primer (CGGACTGGCCGACGCCCC) were annealed at a temperature of 78°C and overhangs were used to introduce point and di-nucleotide mutations through inverse PCR. The

mutated SHOxi constructs were verified by Sanger sequencing (Genewiz). The empty vector pta1300 and the various mutant SHOxi constructs were individually transformed into H53 and H98 Δ*SHOxi H. volcanii* strains as previously described. Overexpression of the mutant SHOxi transcripts were induced under no challenge conditions and harvested for RNA extraction and cDNA generation as described in the overexpression experiments. Malic enzyme mRNA expression was then measured via qPCR (SYBR Green PCR Master Mix, ThermoFisher) in response to mutant SHOxi overexpression using primers for malic enzyme and rp130 as a housekeeping gene.

Dinucleotide luciferase assay. All dinucleotides (NAD^+, NADH, $NADP^+$, NADPH) were extracted using a custom protocol from Promega. After dinucleotide extraction, extracts were used in the corresponding Glo™ Assay (Promega) using the standard protocol where 50 uL of extract was added with 50 uL of Glo™ Detection Reagent (Promega) in a white bottom 96 well plate (Corning). After 30 minutes of incubation the plates were measured for luminescence using a ## luminometer.

Protein carbonyl western blotting. Cells were treated and pelleted as previously described. For protein extraction, frozen pellets were resuspended in 1 mL ice cold 1M salt buffer (50 mM potassium phosphate pH 7.0, 1M NaCl, 1% 2-mercaptoethanol) and sonicated 30 seconds ON/ 30 seconds OFF for 30 minutes. Lysates were centrifuged at 12,000 x g for 30 minutes at 4 C and supernatant was transferred into a new tube and kept on ice. Protein concentrations were measured using the Quick start Bradford 1x assay standard protocol (Bradford). Protein carbonyls were measured by western blotting using the OxyBlot kit standard protocol. Briefly, 20 ug of proteins were derivatized with DNPH and ran on a 4-20% SDS PAGE at 120 V for 30 mins. Protein was transferred to 0.2 uM PVDF membrane

(Ambion) in a Trans-blot Turbo (BioRad) for 7 minutes and incubate with primary and secondary antibodies for 1 hour at room temperature, respectively. ECL+ reagent was added to blots and incubated at room temperature and blots were scanned with a Typhoon phosphoimager.

RNA-seq data. All raw read and processed data from these experiments are available at NCBI under BioProject XXXX. Illumina raw sequence data (.fastq) for each replicate and condition are deposited in NCBI Sequence Read Archive.

CONTRIBUTIONS

DRG: Conceptualization, Investigation, Methodology, Project administration, Writing – original draft, Writing – review & editing

RR: Investigation – sample acquisition, genetics.

KW: Investigation – sample acquisition, genetics.

SD: Investigation – sample acquisition, genetics, qPCR.

JDR: Conceptualization, Funding acquisition, Project administration, Supervision, Validation, Writing – review & editing

ACKNOWLEDGEMENTS

This work was supported by NASA grant 18-EXO18-0091 and Air Force Office of Scientific Research grant FA9550-14-1-0118. We thank Kate Huffer for help in generation of SHOxi deletion mutants. We thank Katie Farney for extensive manuscript editing. We thank David Mohr for sequencing efforts, and Dr. John Kim, Dr. Gisela Storz, Dr. Sarah Woodson, German Uritskiy, and Emine Ertekin for experimental advice and helpful discussions.

SUPPLEMENTAL FIGURES

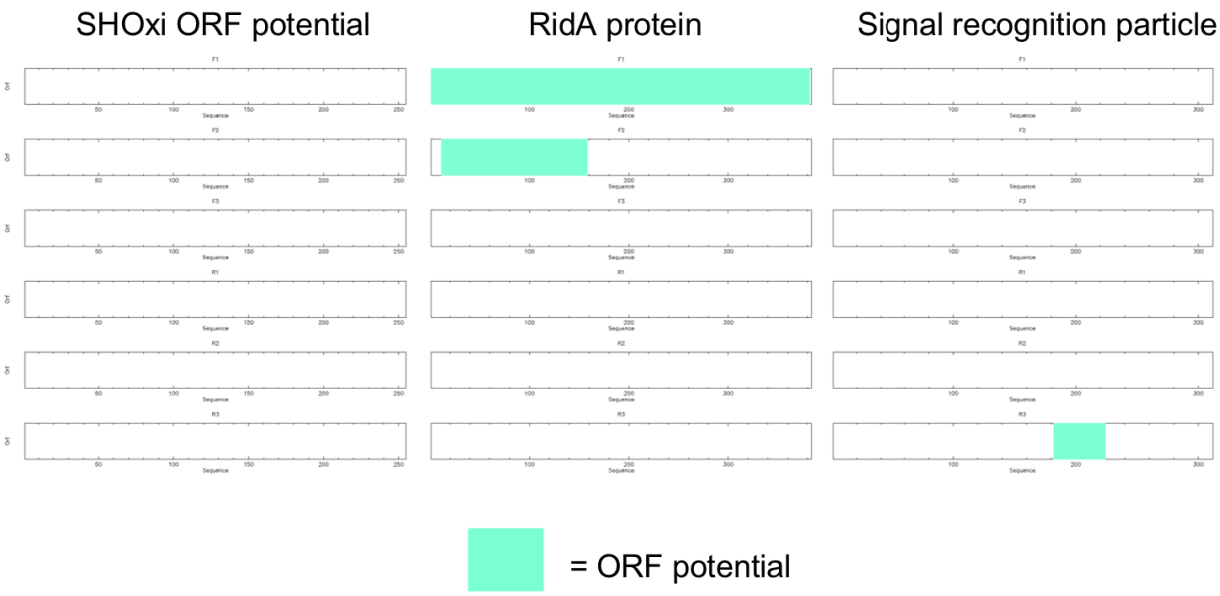

FIGURE 4-S1:

ORF potential plots for the six potential open reading frames of SHOxi, the protein coding mRNA RidA, and the verified non-coding RNA signal recognition particle.

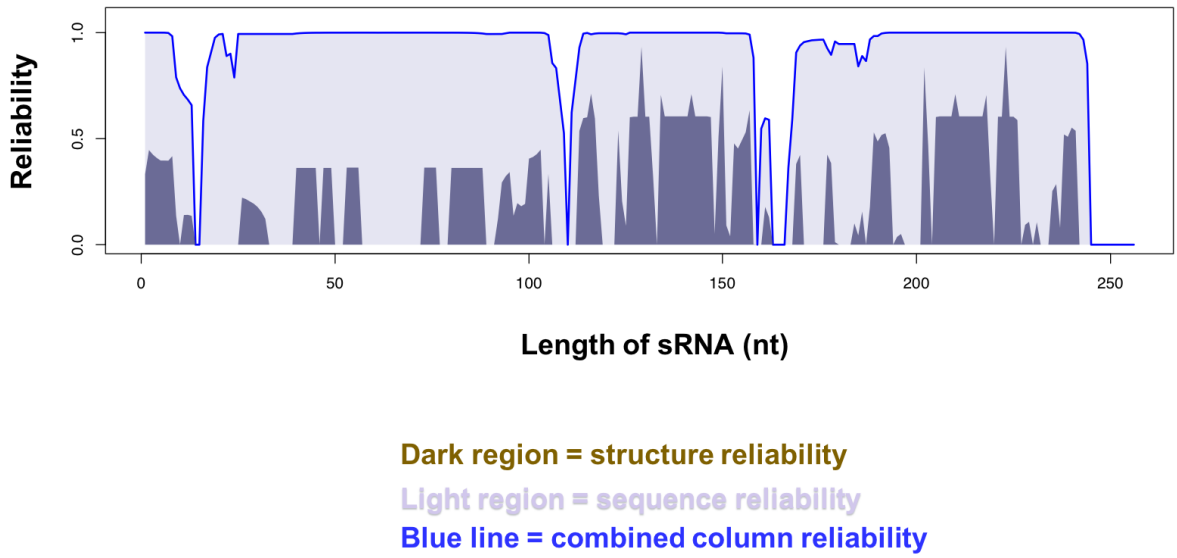

FIGURE 4-S2:

Reliability plot of the secondary structure of SHOxi. Reliability scores are calculated based off of either structure conservation or sequence conservation across the length of the sRNA in the context of its secondary structure.

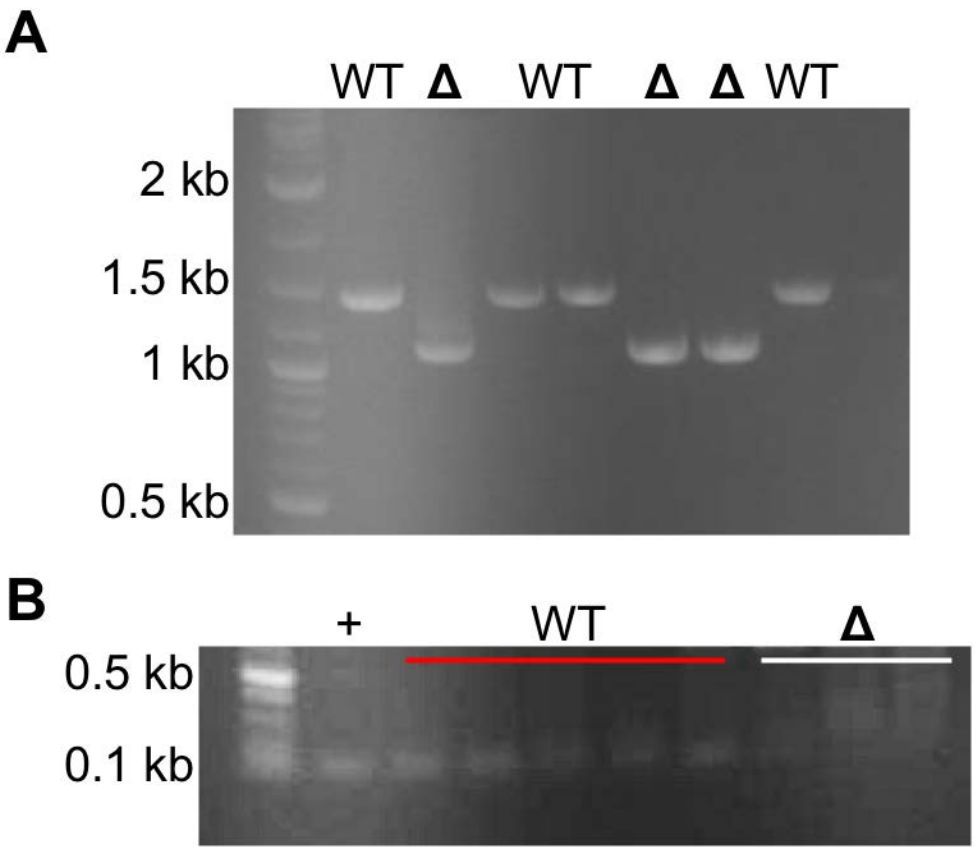

FIGURE 4-S3:

(A) PCR confirmation of SHOxi deletion in the genome of *Hv.* WT denotes wild type genomic DNA and Δ denotes *ΔSHOxi* genomic DNA. Primers targeted 500 bp upstream and downstream of SHOxi and the expected product size was 1.5 kb. (B) RT-PCR confirmation of SHOxi deletion and depleted transcript levels. +, wild type genomic DNA (positive control); WT, wild type cDNA; Δ, *ΔSHOxi* cDNA. Primers were 100 bp internal to SHOxi and the expected product size was 100 bp.

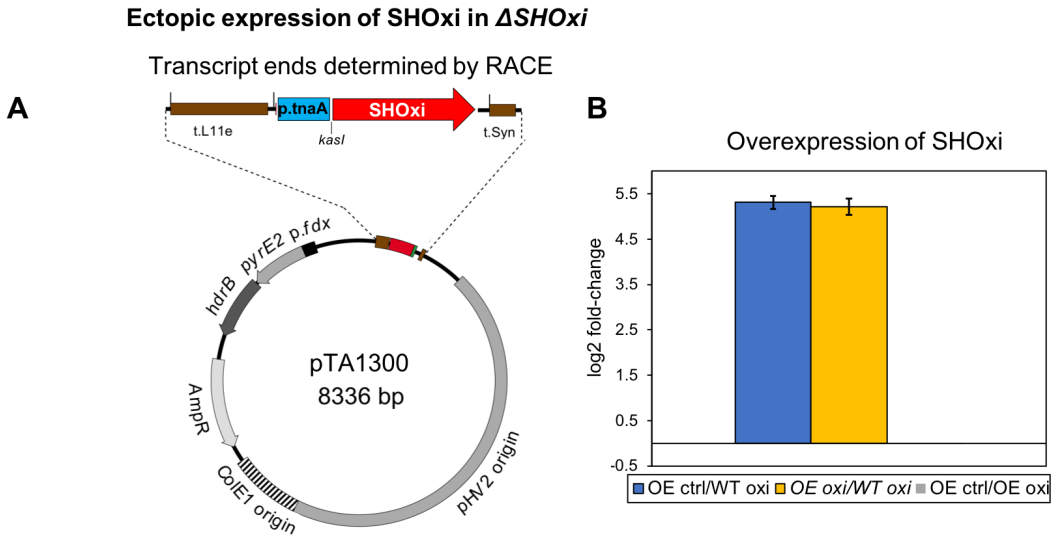

FIGURE 4-S4:

(A) Plasmid map for the newly constructed expression vector pTA1300 for overexpression of SHOxi under an inducible trp promoter. (B) Fold changes of SHOxi overexpression from pTA1300 relative to wild type SHOxi endogenous expression under oxidative stress measured by RNA-seq. OE ctrl is the over expression of wild type SHOxi under no challenge conditions and OE oxi is the over expression of wild type SHOxi under oxidative stress conditions.

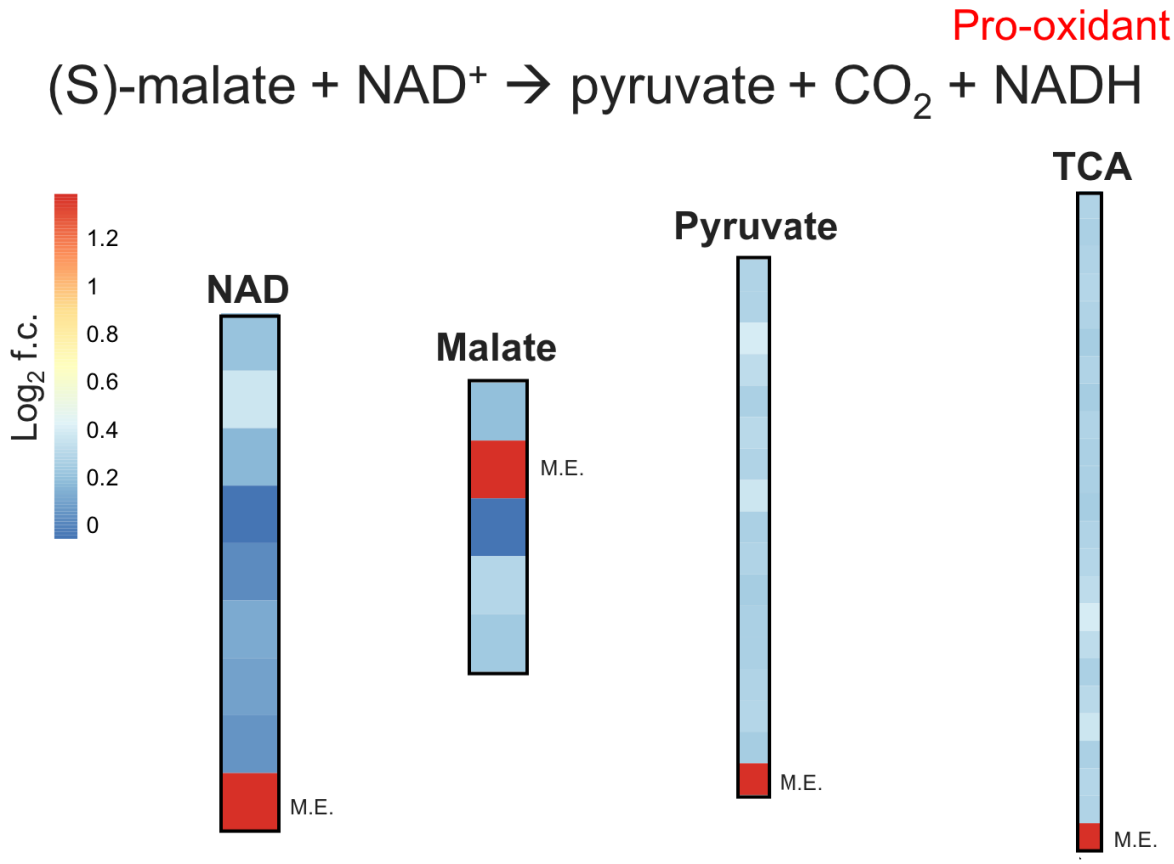

FIGURE 4-S5:

Heatmaps of fold-changes for genes involved in NAD, malate, pyruvate, and TCA metabolism all of which malic enzyme (ME) mRNA plays a role. Fold changes were calculated between wild type and *ΔSHOxi* under oxidative stress.

Chapter 5

Regulatory non-coding small RNAs are diverse and abundant in an extremophilic microbial community

This work has been published in mSystems - Gelsinger DR, Uritskiy G, Reddy R, Munn A, Farney K, DiRuggiero J. 2020. mSystems 5:e00584-19. https://doi .org/10.1128/mSystems.00584-19.

ABSTRACT

Regulatory small RNAs (sRNAs) play large-scale and essential roles in many cellular processes across all domains of life. Microbial sRNAs have been extensively studied in model organisms but very little is known about the dynamics of sRNA synthesis and their roles in the natural environment. In this study, we discovered hundreds of intergenic (itsRNAs) and antisense (asRNAs) sRNAs expressed in an extremophilic microbial community inhabiting halite nodules (salt rocks) in the Atacama Desert. For this, we built SnapT – a new sRNA annotation pipeline that can be applied to any microbial community. We found asRNAs with expression levels negatively correlated with that of their putative overlapping target and itsRNAs that were conserved and significantly differentially expressed between 2 sampling time points. We demonstrated that we could perform target prediction and correlate expression levels between sRNAs and predicted target mRNAs at the community level. Functions of putative mRNA targets reflected the environmental

challenges members of the halite communities were subjected to, including osmotic adjustments to a major rain event and competition for nutrients.

IMPORTANCE Microorganisms in the natural world are found in communities, communicating and interacting with each other, therefore, it is essential that microbial regulatory mechanisms, such as gene regulation effected by sRNAs, be investigated at the community level. This work demonstrates that metatranscriptomic field experiments can link environmental variation with changes in RNA pools and have the potential to provide new insights into environmental sensing and responses in natural microbial communities through non-coding RNA mediated gene regulation.

INTRODUCTION

Non-coding RNAs (ncRNAs) are untranslated short transcripts that are found in the three domains of life where they play essential roles in many cellular processes (Cech & Steitz, 2014; D. Gelsinger & J. DiRuggiero, 2018). In prokaryotes, a subset of these ncRNAs, thereby called small RNAs (sRNAs), is specifically involved in gene regulation through RNA-RNA mediated interactions, modulating core metabolic functions and stress-related responses (Gottesman & Storz, 2011). These sRNAs range from 50 to 500 nucleotides in size and can be of two types: *trans*-encoded sRNAs, also called intergenic sRNAs (itsRNAs), which bind their mRNA targets via imperfect base-pairing and can target multiple genes, including key transcription factors and regulators (Wagner & Romby, 2015). itsRNAs can activate or inhibit translation initiation by interacting with the ribosome binding site (RBS) and/or modulating mRNA stability (Wagner & Romby, 2015). In contrast, *cis*-encoded antisense RNAs (asRNAs) are transcribed on the DNA

strand opposite their target gene and thus can act via extensive base pairing; they have been found to repress transposons and toxic protein synthesis (Wagner & Romby, 2015).

The functional roles of microbial sRNAs have been extensively studied in a few model organisms and very little is known about the dynamics of sRNA synthesis in natural environments and the roles of these short transcripts at the community level (Carrier, Lalaouna, & Massé, 2018; D. Gelsinger & J. DiRuggiero, 2018). To our knowledge, only a handful of studies have reported the discovery of sRNAs in natural microbial communities (Bao, Wang, Doak, & Ye, 2015; Duran-Pinedo, Yost, & Frias-Lopez, 2015; Hou, Pfreundt, Miller, Berman-Frank, & Hess, 2016; Shi, Tyson, & DeLong, 2009) and there is no publicly-available bioinformatic tool for sRNA discovery in single-species isolates and in the metagenomic context (D. Gelsinger & J. DiRuggiero, 2018). This paucity of knowledge suggests that an abundance of sRNAs remain to be discovered with potentially essential roles in stress response (Clouet-d'Orval et al., 2018), inter-species communication, and/or cross-species RNA interference (Cai et al., 2018; Toyofuku, Nomura, & Eberl, 2019; Tsatsaronis, Franch-Arroyo, Resch, & Charpentier, 2018). This might be relevant to extreme environments where microbial communities are specifically adapted to a narrow set of environmental conditions, i.e. high salt, low pH and particularly sensitive to perturbations (Gherman Uritskiy et al., 2019).

In hyper-arid deserts, microbial communities find refuge inside rocks as a survival strategy against the extreme conditions of their environment (Pointing & Belnap, 2012). Such community inhabits halite (salt) nodules in Salars of the Atacama Desert, Chile, which is one of the oldest and driest deserts on Earth (Crits-Christoph et al., 2016; Finstad et al., 2017). The halite endolithic (within rock) community harbors mostly members of

the Archaea (*Halobacteria*), unique *Cyanobacteria*, diverse heterotrophic bacteria, and a novel type of algae (Crits-Christoph et al., 2016; Finstad et al., 2017), all of which were shown to be transcriptionally active (G. Uritskiy et al., 2019). The main source of liquid water for this community is from salt deliquescence (Davila et al., 2008) and it is sustained by CO_2 fixed via photosynthesis (Crits-Christoph et al., 2016; Davila et al., 2015). While previous studies have demonstrated the role of sRNAs in the stress response of one of the members of this community, the halophilic archaeon *Haloferax volcanii* (D. R. Gelsinger & J. DiRuggiero, 2018; Kliemt, Jaschinski, & Soppa, 2019), there is no information on any of the other members.

Here we used a combination of genome-resolved metagenomics and metatranscriptomics to investigate the role of sRNAs in the adaptive response of microorganisms inhabiting halite nodules. We developed an analytical pipeline, SnapT, built on our previous work on sRNAs from the archaeon *Haloferax volcanii* (Gelsinger & DiRuggiero, 2018), to enable the discovery of sRNAs at the community level. Using strand-specific metatranscriptomics, we found hundreds of sRNAs (*both* itsRNAs and asRNAs) from multiple trophic levels in the halite community, including conserved sRNAs, validating our experimental approach. Previous studies were limited to either intergenic or antisense sRNAs, never both; analysis of both types of sRNAs in our study allowed for the most comprehensive view of the sRNA regulatory landscape in a microbial community (Bao et al., 2015; Duran-Pinedo et al., 2015; Hou et al., 2016; Shi et al., 2009). A number of itsRNAs were significantly differentially regulated between 2 sampling time points, providing validation that sRNAs can be modulated in the natural environment. For a subset of these, we were able to perform structure and target prediction of conserved

sRNAs to decipher their potential regulatory roles, a first at the metatranscriptomic level. Coupling metagenomics and metatranscriptomics with SnapT allows for the potential to uncover the complex regulatory networks that govern the state of a microbial community.

RESULTS

Landscape of predicted sRNAs in the halite community and validation. We discovered hundreds of ncRNAs in an extremophilic community inhabiting halite nodules (salt rocks) in the Atacama Desert by using SnapT (https://github.com/ursky/SnapT), a bioinformatic tool for sRNA discovery (**Table 2**). We used metatranscriptomics data from multiple replicate samples collected in the field in 2016 and 2017 (21 and 24 replicates for 2016 and 2017, respectively; Table S1). Using SnapT, we aligned reads from stranded RNA-seq libraries to our reference co-assembled metagenome from a previous study (Gherman Uritskiy et al., 2019) (**Fig. 5-S1**). The assembled transcripts were then intersected with the metagenome annotation as well as open reading frames to select for either novel transcripts on the opposite strand of coding transcripts (asRNAs) or for novel transcripts that fell into intergenic regions (itsRNAs). Putative ncRNA transcripts were then further enriched using a threshold at 5x and 10x assembly coverage to identify intergenic and antisense ncRNAs, respectively. (**Fig. 5-S2A; Table 2).** The size of these ncRNAs was then filtered from 50 to 500 nucleotides to produce a final set of non-coding sRNAs. The size distribution of these sRNAs was primarily between 50 and 200 nt for itsRNAs and above 200 nt for asRNAs. (**Figs. 5-S2B and 5-S2C**).

Table 2: Summary of ncRNAs discovered in halite community

	Number (%)*	% in Archaea	% in Bacteria
Total ncRNAs	1538 (100)	54	46
Rfam ncRNAs	79 (5)	73	27
Conserved sRNAs**	155 (10)	60	40
Antisense sRNAs	925 (60)	40	60
Intergenic sRNAs	613 (40)	75	25

*Percent from total ncRNAs

** Conserved other than Rfam ncRNAs

The halite ncRNAs were taxonomically assigned to diverse members of the community; their distribution between Archaea (54%) and Bacteria (46%) (**Table 2**) were similar to that of the total metatranscriptomic reads for the community (**Fig. 5-1B-C**). In contrast, the taxonomic profile of the metagenome showed a larger contribution of bacterial reads and in particular of reads assigned to *Cyanobacteria* and *Bacteroidetes* (**Fig. 5-1A**). Because of the use of strand-specific RNA-seq libraries, we could confidently identify both intergenic (it)sRNA, located between coding regions, and antisense (a)sRNA, overlapping with their putative target (**Table 2**).

We found 3 times more itsRNAs in the Archaea than in the Bacteria, whereas asRNAs were more abundant in the Bacteria and more often associated with members of the *Cyanobacteria* (38%) and *Bacteroidetes* (15%) (**Table 2**; **Fig. 5-1D-E**). We also found

79 ncRNAs, that belong to 6 known families of RNAs present in the Rfam database (**Fig. 5-S2D**) (Kalvari et al., 2018), validating our experimental and computational approach. This database is a collection of RNA families, each represented by multiple sequence alignments, consensus secondary structures, and covariance models. Of the Rfam-conserved ncRNAs, 70% were assigned to archaea and included RNaseP RNAs, signal recognition particle RNAs (SRP RNAs), and tRNAs. Of the Rfam-conserved bacterial ncRNAs, most were from SRP RNAs and tRNA conserved families. In addition, a cobalamin riboswitch and the regulatory sRNA, CyVA-1, were detected in low abundance in the halite *Cyanobacteria*. We also found 3 ncRNAs (4%) from Eukarya, a tRNA, a U4 spliceosomal RNA, and a RNase for mitochondrial RNA processing (MRP). Using blastn analysis (max e-value of 1E-3, sequence similarity of 70% or more, coverage of 50% or more), we discovered another 155 ncRNAs that were conserved in the NCBI nt database, with 60% from archaea and 40% from bacteria (**Table 2**). The majority were asRNAs (109), with only 44 itsRNAs. The conserved asRNAs most highly expressed (standardized tpm> 100) were all SPR RNAs in haloarchaea that were not found in the Rfam database. Of the conserved itsRNAs, we identified 3 tRNAs, 13 SRP RNAs, and 22 ncRNAs that were found in the genome of multiple species, all *Halobacteria*, but with no function assigned. The most highly expressed and conserved itsRNAs (standardized tpm> 100; 13

ncRNAs) were SRP RNAs not included in the Rfam database.

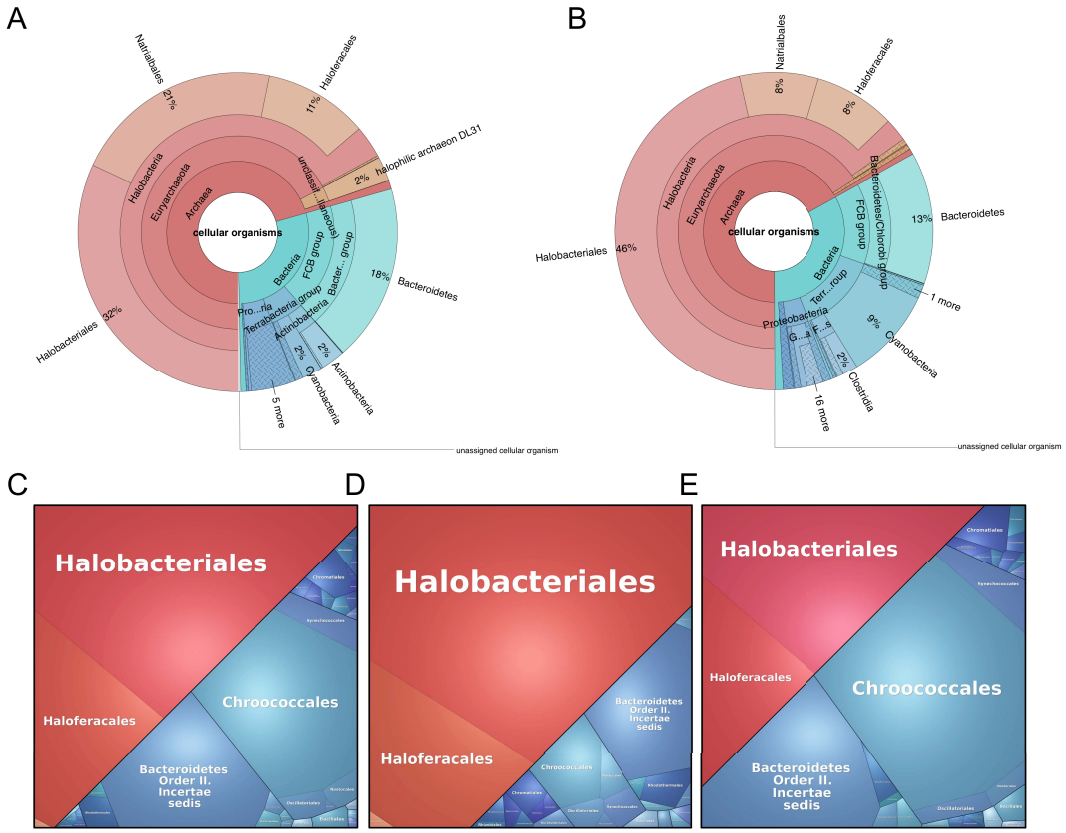

Figure 5-1: Taxonomic distribution. Krona graphs of (A) the halite metagenome based of DNA sequence reads and (B) the halite metatranscriptome based on RNA sequence reads; and Voronoi plots of (C) total sRNAs; (D) itsRNAs and (E) asRNAs discovered in the halite community. Partitions of the Voronoi plots correspond to relative abundances of the indicated taxa.

Another validation of our findings was the presence of canonical promoter elements upstream of archaeal itsRNAs, suggesting that they were indeed *bona fide* transcripts that could recruit basal transcription factors (**Fig. 5-S3A).** We did not find significant promoter elements upstream of the bacterial itsRNAs, which might reflect the diversity of promoter elements across the various bacterial taxa we identified in the halite community. In contrast, no promoter elements were identified in the upstream regions of asRNA from

both domains of life.

When looking at the expression levels of all itsRNAs normalized to contig abundances, we found that they were similar for both the 2016 and 2017 samples and slightly higher than that of the asRNAs, whereas the expression profile of the asRNAs was more variable across samples for both years (**Fig. 5-S3B**). Remarkably, the expression levels of itsRNAs and asRNAs for both years was 2-fold higher than that of protein encoding genes. Whereas there is an inherent bias in our approach to identifying sRNAs at the community level (coverage threshold in SnapT) compared to protein-encoding genes, this finding strongly indicates potential functional relevance for a number of these sRNAs.

We experimentally validated several sRNAs using RT-PCR with environmental and enrichment cultures. Enrichments were performed with several media containing high (25%) and relatively low (18%) salt, and various combinations of carbon sources. Amplicon sequencing of the enrichments revealed that high salt and diverse carbon sources resulted in a higher diversity of taxa, although haloarchaea dominated in all enrichments (**Fig. 5-S4**). All validated sRNAs belong to haloarchaea except for one from *Cyanobacteria*. Sequences of the PCR products confirmed that they were sRNAs and validated our computational approach.

Relationship with target genes and putative function of community asRNAs. Using our strand-specific RNA-seq data, we were able to identify the overlap position of asRNAs to their antisense transcripts. We found that, in both Archaea and Bacteria, the majority of asRNAs start within the span of their cognate gene and end near the 5' end of its mRNA. In both domains, there is also an enrichment for asRNA-mRNA overlaps near

the 5' end of the mRNA (**Fig. 5-S2E**). A similar trend has previously been reported in two species of archaea (de Almeida et al., 2019; D. R. Gelsinger & J. DiRuggiero, 2018).

We compared the expression level of asRNAs with that of their putative target genes and found that highly expressed asRNAs were associated with lowly expressed genes (**Fig. 5-2A**). Of gene pairs with asRNA expression >100 tpm and gene expression <0.1 tpm, most where from haloarchaea (77%), with 12% of *Cyanobacteria*, and 11% of other bacteria (*Bacteroidetes* and A*cinetobacter*). Gene functions were enriched for transport (16%) and cell membrane/wall metabolism (5%), while most were hypothetical proteins (44%). Of the genes potentially negatively regulated by their cognate asRNAs, we found an archaeal regulator of the IclR family and potassium uptake protein TrkA. Only 2 asRNAs with high expression levels (>100 standardized tpm) were associated with genes with relatively high expression levels (>1 standardized tpm), while still being negatively correlated **(Fig. 5-2A)**. The corresponding genes encoded for an iron complex outer-membrane receptor protein from *Salinibacter* and an ABC-type sodium efflux pump permease subunit from an *Halobacteria*. When applying a stringent cut-off, we found 9 statistically significant and negatively correlated asRNA:gene pairs **(Fig. 5-2B** and **Fig 5-S5A).** Four were from *Bacteroidetes*, 4 from *Halobacteria*, and 1 from an unidentified bacterium. At the functional level, transport systems, and in particular iron transport systems, were particularly enriched. In contrast, we did not find any significant positive regulation between asRNAs and their cognate genes. When adjusted for the carrying organism's abundance, expressed as the average RNA read coverage of the contigs, we found that overall itsRNAs were more highly expressed than asRNAs (**Fig. 5-2C-D**).

Highly expressed sRNAs, for both types, were mostly carried by haloarchaea.

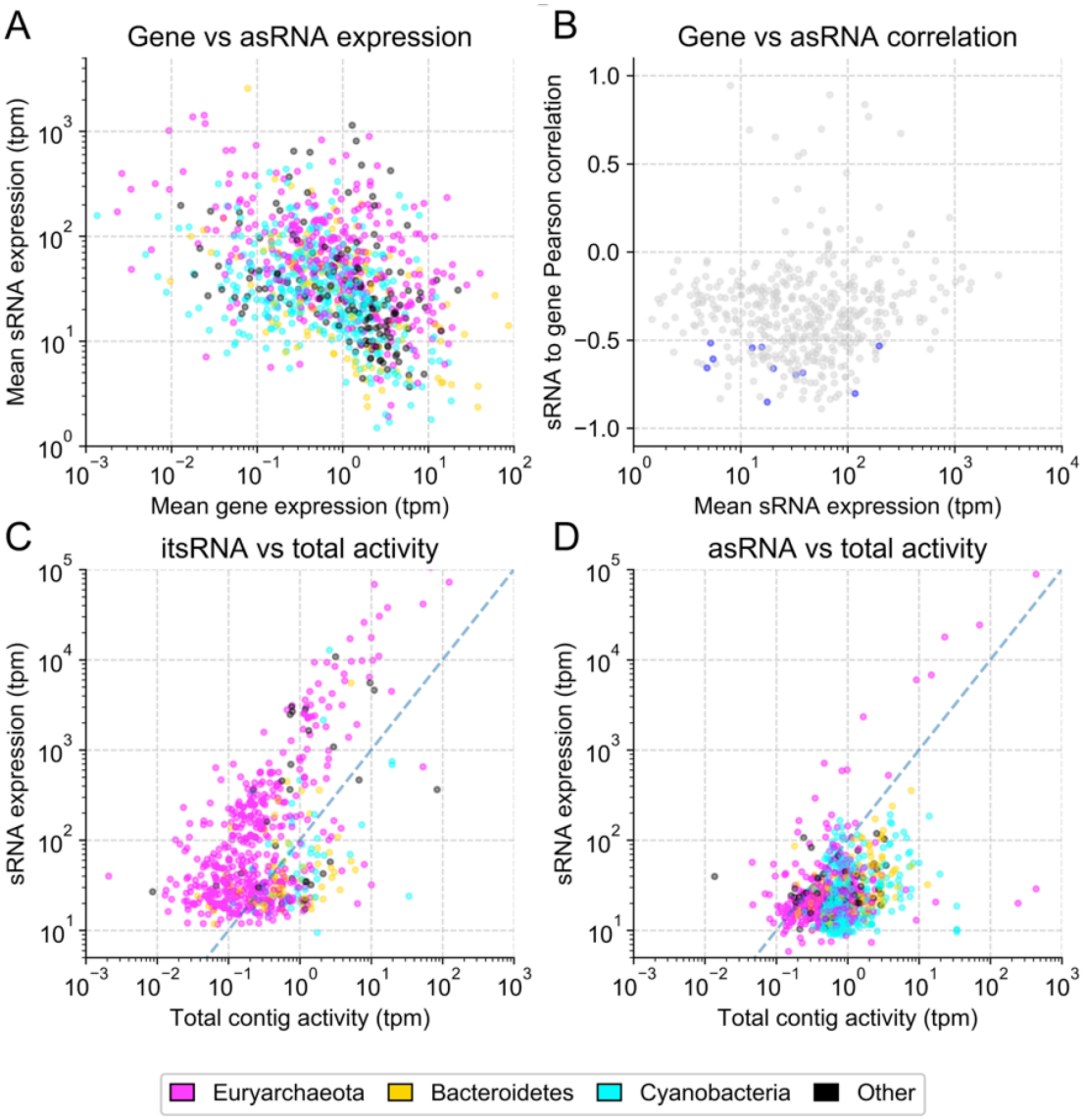

Figure 5-2: RNA expression levels. (A) asRNAs and their putative targets (mean expression levels of all replicates) (TPM); (B) Pearson correlations in expression level between asRNAs and their putative mRNA targets across all the replicates, with significant correlations (pval<0.01) highlighted in blue; (C) average expression of itsRNA and average expression of (D) asRNAs over the average expression of the contigs on which they are found. Dashed lines are added for simpler visual interpretation and a represent 1:1 ratio of contig activity to sRNA expression. Figure generated by GU.

Differential expression of itsRNAs at the community level and target prediction. Analysis of itsRNAs expression levels showed a clear separation between the 2016 and 2017 samples (**Fig. 5-3A**), which was confirmed by the analysis of metatranscriptomic expression levels of annotated genes from the metagenome (**Fig. 5-S5B**). We carried out a differential expression analysis and found that 109 (18%) of the regulatory itsRNAs were significantly differentially expressed (FDR <5%) between samples collected in 2016 and 2017 (**Fig. 5-3B**), 3 and 15 months after a major rain event in the desert (Gherman Uritskiy et al., 2019). Of these, 72% were annotated as archaea and 28 % as bacteria and 16 were conserved in multiple genomes (14 from *Halobacteria* and 2 from *Cyanobacteria*). Conservation of differentially expressed itsRNAs allowed for structure modeling and, when high-quality MAGs (>70% completion and <5% contamination) were available from the metagenome, target prediction (**Fig. 5-4 and Fig. 5-S6A**). Several non-differentially expressed itsRNAs were also conserved, providing additional opportunity for structure prediction; these included itsRNAs from *Halococcus* (STRG. 48671.1; 69 nt), *Halobellus limi* (STRG136887.1; 209 nt), and from a member of the *Nanohaloarchaea (*STRG.4577.1; 266 nt) (**Fig. 5-S6A).**

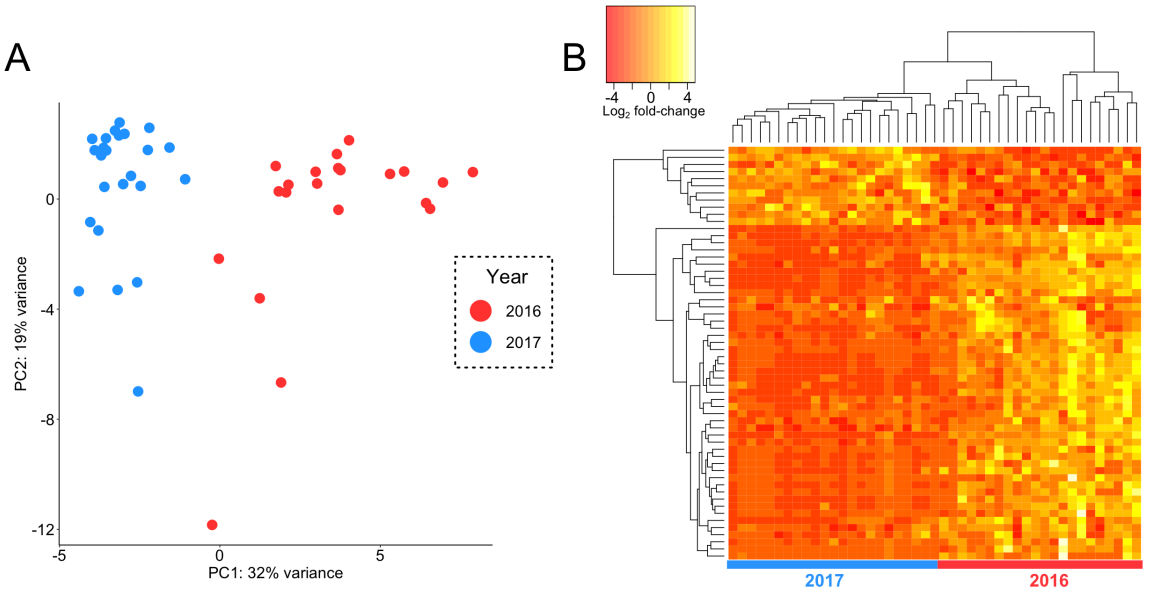

Figure 5-3: itsRNA differential expression. (A) PCA plot showing itsRNA expression levels clustered by year and (B) heat map of log_2-transformed fold change for the top 50 significantly differentially expressed itsRNAs; each row is an itsRNA and each column a sample collected in 2016 or 2017.

All predicted structures displayed loop and stems regions that had high sequence conservation (light purple regions on sequence–structure-based alignment reliability [STAR] profile plots) and high structure conservation (dark purple), and line plots representing the reliability of the predictions as calculated by LocaRNA (**Fig. 5-4 and Fig. 5-S6B**). Density plots combined with dumbbell plots were used for visualizing predicted interactions between itsRNAs and their putative targets, using IntaRNA data from the top 100 most reliable interaction predictions with the lowest free energy of hybridization (Mann et al., 2017) **(Fig. 5-4)**. High confidence assignments were obtained for 4 differentially expressed itsRNAs from *Cyanobacteria*, *Halapricum salinun*, and a member of the *Halobacteria*. More than one interaction peak were derived from density plots; peak

1 (green) corresponded to the highest interaction density, which mapped to loop regions in the itsRNA secondary structure with high sequence and structure conservation, respectively, and was thus a confident assignment as an interaction region, whereas Peak 2 (yellow) was a less confident assignment structurally despite high interaction density (**Fig. 5-4 and Fig. 5-S6B**).

Using this information, we identified the most probable targets for *Cyanobacteria* STRG.5354.4 candidate itsRNA (229 nt). This itsRNA was conserved as a 6S regulatory RNA in the rfam database, which in bacteria is found to inhibit transcription by binding directly to the housekeeping holoenzyme form of RNA polymerase (Wassarman, 2018). Of the top 50 most probable targets for STRG.5354.4, which were those with the lowest free energy of hybridization between itsRNA and targets, were cation/H+ antiporters [shown to be involved in osmoregulation (Krulwich, Hicks, & Ito, 2009), a PleD family two-component response regulator, the photosystem I PsaB protein, chemotaxis transducers, and proteins involved in energy metabolism. Most probable targets for differentially expressed itsRNA, STRG. 86294.1 (281 nt) from *Halapricum salinum* included various transporters and putative membrane and cell wall-associated proteins; notably an ammonium transporter (Amt family), an alkanesulfonate monooxygenase SsuD from a gene cluster expressed under sulfate or cysteine starvation (Eichhorn, van der Ploeg, & Leisinger, 1999)**,** and several proteins involved in cofactors and vitamin metabolism. Predicted targets with the lowest free energy of hybridization for STRG.49508.3 candidate itsRNA (99 nt) from *Halobacteria* were elongation factor 1-alpha, which promotes the GTP-dependent binding of aminoacyl-tRNA to the A-site of ribosomes during protein biosynthesis, several ribosomal proteins and hypothetical proteins. Target prediction for

Cyanobacteria STRG.5356.1 candidate itsRNA (242 nt) included molecular chaperones (DnaK and DnaJ classes), a cell division protease FtsH, and several uncharacterized proteins.

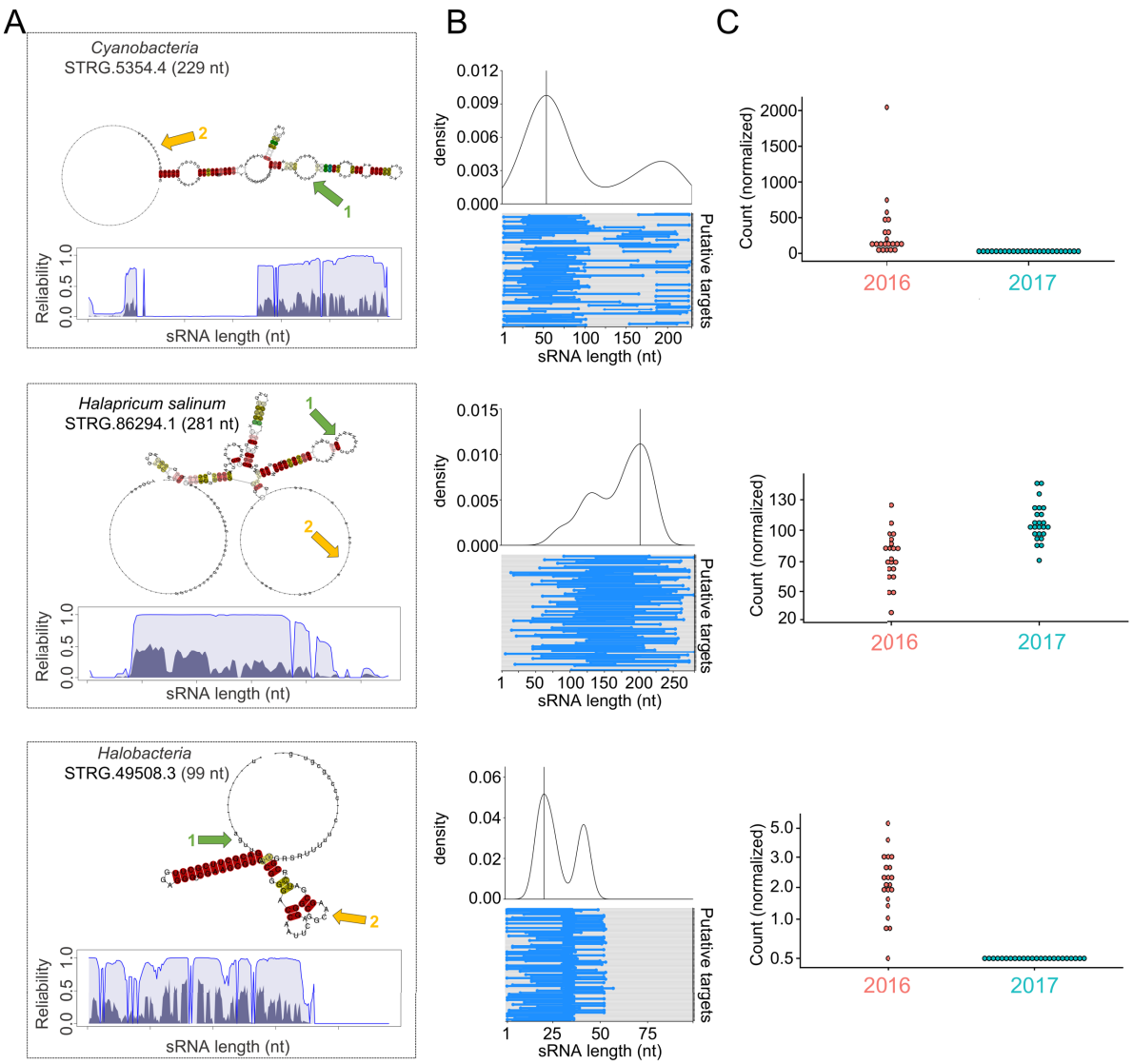

Figure 5-4: Predicted structure, target identification, and expression levels for selected differentially expressed itsRNAs. (A) 2D-layout of consensus structures with base pairs coloring showing sequence and structure conservation and interactions peaks (green and yellow arrows); STAR profile plots with dark regions indicating structure reliability, light regions representing sequence reliability, and thin lines showing the combined column-reliability as computed by LocARNA-P. Structures are using minimum free energies (MFE). (B) Interaction plots of itsRNAs and their predicted targets. The top graphs are density plots calculated from the top 100 putative targets, and on the bottom are dumbbell plots of interactions (blue dumbbells) along the length of the itsRNA for the top 100 predicted mRNA targets; interaction peaks are shown in green and yellow in the predicted structures; (C) Expression levels represented as normalized count for each

DISCUSSION

The roles of regulatory sRNAs have been extensively studied in bacteria, and to a lesser extent, in archaeal model systems (Carrier et al., 2018; D. Gelsinger & J. DiRuggiero, 2018) but, to date, only four studies have reported the discovery of sRNAs in microbial communities. In one study, Shi *et al*. (Shi et al., 2009) used metatranscriptomic data to identify unique microbial intergenic sRNAs in the ocean's water column, in a second study, Bao et al. (Bao et al., 2015) revealed extensive antisense transcription in the human gut microbiota, also using metatranscriptomic datasets. In the last two studies, Hou et al. (Hou et al., 2016) conducted a survey of transcription start sites (TSS) and identified a small number sRNA TSS, while in Duran-Pinedo et al. (Duran-Pinedo et al., 2015) carried out an extensive study of intergenic sRNAs 45 m deep in the northern Red Sea and focused on those that were conserved in the rfam database. Efforts have also been made to mine

publicly available databases for sRNA discovery (Weinberg et al., 2017) but this was still addressing the role of sRNAs in single microorganisms. Each of these studies was limited to one type of sRNA (intergenic or antisense) usually due to technical limitations (i.e. sequencing technology, library preparations). Through the combination of strand-specific RNA-sequencing and the development of the first microbial sRNA identification pipeline, SnapT, we were able to comprehensively identify all sRNAs in an extremophilic microbial community. This combination of technologies allowed for a highly resolved view of sRNA-mediated regulation from multiple trophic levels in the community, from primary producing cyanobacteria to the dominant heterotrophic haloarchaea.

One major difficulty in obtaining metatranscriptomic data from natural microbial communities, in particular from extreme environments, is the low amount of biomass that can be collected, resulting in low RNA yields (Gherman Uritskiy et al., 2019). This, in turn, prevents attempts at ribo-depletion, resulting in a decreased number of non-ribosomal RNA reads available for analysis. Nevertheless, using SnapT, a flexible pipeline to process metagenomics and metatranscriptomic data, we report the discovery of hundreds of diverse sRNAs from an extremophilic community inhabiting halite nodules in the Atacama Desert. In the process, we applied extensive quality control with coverage thresholding, correction for contig edge misannotation, and the removal of potential non-ncRNAs through sequence and homology searches. While this approach might potentially result in false negatives and may bias our findings toward the most highly expressed sRNAs in the community, it also insured the robustness of our sRNA predictions by minimizing the number of false positives. The identification of ncRNAs in the halite community that belongs to the Rfam database (Kalvari et al., 2018), together with experimental validation of several sRNAs

with environmental and enrichment cultures, substantiated our analytical approach. Additionally, expression levels of sRNAs 2-fold higher than that of protein-encoding genes, strongly indicate potential functional relevance for a number of these sRNAs.

The taxonomic composition of the halite sRNAs matched that of the community's metatranscriptomic profile, reflecting the contribution of the most active members, including *Cyanobacteria*, *Bacteroidetes*, and several *Halobacteria*. We found significantly more itsRNAs in the archaea than in the bacteria and the trend was reversed for the asRNAs. This novel finding is representative of published work in model organisms where a wide range of sRNAs has been found so far in prokaryotes, from less than a dozen to more than a thousand per genome (**Fig. 5-S7**) (Carrier et al., 2018; Gelsinger & DiRuggiero, 2018).

Antisense sRNAs overlap their putative targets providing insights into their functional role (Wagner & Romby, 2015). In the halite community, we found that asRNAs expression levels were negatively correlated with that of their putative targets, with highly expressed asRNAs overlapping lowly expressed protein-encoding genes. A similar trend was reported in the haloarchaeon *Haloferax volcanii*, when investigating oxidative stress-responsive sRNAs, and most of the putative targets were transposase genes (Gelsinger & DiRuggiero, 2018). Putative target gene functions in our study were mostly from haloarchaea and enriched for transport systems, cell membrane and cell wall metabolism, with a large number of hypotheticals. Of particular interest, was an archaeal IcIR transcription regulator; these regulators are known to be involved in diverse physiological functions, including multidrug resistance, degradation of aromatics, and secondary metabolites production (Molina-Henares, Krell, Eugenia Guazzaroni, Segura, & Ramos,

2006) and are distributed in a wide range of prokaryotes, including Archaea (Perez-Rueda et al., 2018). Also of interest, was a Trk potassium uptake system, also found in both bacteria and archaea and essential for the maintenance of high intracellular potassium in salt-in strategists (Oren, 2013). Salt-in strategists accumulate KCl to balance the high osmotic pressure of their environment, hence the need to actively pump potassium in the cell. In contrast, we did not find any significant positive regulation between asRNAs and their cognate genes (up-regulation of both), which might be due to the inherent quality of our data set, i.e. no ribo-depletion and heterogeneity across replicates (Gherman Uritskiy et al., 2019). Alternatively, it might also reflect promiscuous transcription processes as argued when considering the functionality of asRNAs (Lloréns-Rico et al., 2016). Other arguments in favor of spurious transcription were the size distribution for asRNAs found in the halite community, which was significantly larger than that of itsRNAs, low expression level when adjusted for organism abundance when compared to itsRNAs, and the absence of canonical regulatory elements in the upstream regions of asRNAs. However, we found also putative target functions that reflected the environmental challenges faced by members of this extremophile community, such as osmoregulation and nutrient uptake, indicating that these asRNAs might indeed regulate fundamental biological functions at the community level.

We previously showed that the halite community dramatically shifted its taxonomic and functional composition after a major rain event in 2015, and while it recovered at the functional level in 2017, 15 months after the rain, members of the communities were permanently replaced (Gherman Uritskiy et al., 2019). Here we found that 18% of the halite community itsRNAs were significantly differentially expressed (FDR <5%) between

samples collected in 2016 and 2017 (3 and 15 months after the rain, respectively), potentially indicating a transcriptional response to changes in environmental conditions. Intergenic sRNAs are of particular interest because they can target multiple genes, including key transcription factors and regulators (Gottesman & Storz, 2011). As a consequence, a single sRNA can modulate the expression of large regulons and thus have a significant effect on metabolic processes (Carrier et al., 2018). However, they do not overlap their target genes or bind their target mRNAs with perfect complementary, which makes finding targets for these sRNAs very challenging without genetic tools (D. Gelsinger & J. DiRuggiero, 2018).

To solve this problem at the community level, we focused on itsRNAs that were conserved and for which we could perform structural prediction. The intersection of this small subset of sRNAs with high-quality MAGs that could be used as reference genomes, yielded confident target predictions for 4 differentially expressed itsRNAs, giving insights into metabolic functions potentially regulated by sRNAs at the community level. These included transporters, particularly related to osmotic stress, nutrient uptake and starvation, and pathways for chemotaxis and energy production and conversion. These pathways reflect the environmental challenges members of the halite communities are subjected to, including osmotic adjustments to climate perturbation (Gherman Uritskiy et al., 2019) and competition for nutrients in a near-close system with primary production as the major source of organic carbon (Crits-Christoph et al., 2016). Using the genomic context of sRNAs from the ocean's water column microbial communities, Shi *et al*. (Shi et al., 2009) reported similar metabolic functions, underlying specific regulatory needs for natural communities. In contrast, genes with antisense transcription to asRNAs identified in the

human gut microbiome were mostly transposase genes with a small component of bacterial house-keeping genes (Bao et al., 2015). It important to note that no computational target prediction, using sRNA conserved predicted structure, was reported in either study. Our ability to predict *de novo* targets for sRNAs, drastically increases the scale of regulatory potential we can map to a microbial community. Target prediction is entirely reliant on high-quality MAGs and gene annotation, which we have successfully done through method development (Gherman Uritskiy et al., 2019). Taking this together, we suggest that extremophilic communities, including the halite communities, can be used as model systems to study sRNA dynamics in a natural environment.

Regulation of transcription by 6S sRNA has been shown to increase competitiveness and long-term survival in bacteria (Wassarman, 2018), suggesting an important role for *Cyanobacteria* candidate sRNA STRG.5354.4, identified as a 6S sRNA. Because of high RNA-seq coverage of the *Cyanobacteria* MAGs, we could show that 40% of the top 50 targets for sRNA STRG.5354.4 were differentially regulated and more highly expressed in 2016, suggesting positive regulation by this sRNAs onto its putative targets. Transcriptional factors and regulators were also found as putative targets of differentially regulated itsRNAs in the halite community, underlying the capacity of microbial sRNAs to modulate the expression of large regulons (D. Gelsinger & J. DiRuggiero, 2018; Gottesman & Storz, 2011; Nitzan, Rehani, & Margalit, 2017). Finally, a candidate itsRNAs from the *Halobacteria* had several predicted targets associated with ribosomal proteins and proteins involved in translation processes. This finding, together with a recent study in *H. volcanii* (Wyss, Waser, Gebetsberger, Zywicki, & Polacek, 2018), supports the idea of sRNA modulation of protein biosynthesis in the Archaea. A potential framework for

mechanisms for sRNA regulation of translation might be provided by a report, in the haloarchaeon *Halobacterium salinarum*, of modular translation subsystems that might selectively translate a subset of the transcriptome under specific growth conditions (Raman et al., 2018).

In this study, we characterized the taxonomic and functional landscape of sRNAs across two domains of life in an extremophilic microbial community, demonstrating that asRNAs and itsRNAs can be reliably identified from natural environmental communities. This is essential because sRNAs play essential roles in gene regulation across the 3 domains of life, but most sRNA studies have only been conducted with single organisms. Microorganisms do not live by themselves in the natural environment, they are found in communities and, if we want to understand the molecular mechanisms underlying community stress responses, it is essential to address the role of sRNAs in those regulatory processes. To facilitate this work, we built a flexible pipeline, SnapT (https://github.com/ursky/SnapT), leveraged by our expertise of sRNA biology in a model halophilic archaeon, and which is available to use with metatranscriptomic data from any community. We demonstrated that we could perform target prediction and correlate expression levels between itsRNAs and predicted target mRNAs, paving the way for novel discoveries that have never been done at the community level. While additional work with enrichment cultures remains to be done to fully characterize the functional roles of sRNAs from the halite community, and their mechanism of action, these differentially expressed sRNAs for which we found putative targets show the power of community-level, culture-independent approach analysis for gene regulation processes.

MATERIAL AND METHODS

Sample and weather data collection and nucleic acid extraction. Halite nodules were harvested in Salar Grande, an ancient evaporated lake in the Northern part of the Atacama Desert (Robinson et al., 2015) in February 2016 and 2017, 3 and 15 months after a major rain event (Gherman Uritskiy et al., 2019). All nodules were harvested within a $50m^2$ area as previously described (Robinson et al., 2015). The colonization zone of each nodule was grounded into a powder, pooling from 1-3 nodules until sufficient material was collected, and stored in the dark in dry conditions until DNA extraction in the lab. Samples used for RNA were stored in *RNAlater* at 4°C until RNA extraction in the lab within 14 days of collection. Genomic DNA was extracted with the DNAeasy PowerSoil DNA extraction kit (QIAGEN) as previously described (Crits-Christoph et al., 2016; Robinson et al., 2015) (QIAGEN). Total RNA was extracted from the fixed samples by first isolating the cells through gradual dissolving of the salt particles as previously described (Crits-Christoph et al., 2016; Robinson et al., 2015) and lysing them through mechanical bead beating with the RNAeasy PowerSoil RNA extraction kit (QIAGEN). Total RNA was then extracted from the lysate with a Quick-RNA miniprep kit (RNA >17 nt) (Zymo Research). We obtained 10-100 ng of RNA / g of grounded halite. RT-PCR was used to validate the absence of contaminating DNA in the total RNA used for RNA-seq libraries with 16S rRNA primers 515F/926R (G. Uritskiy et al., 2019).

Library preparation. Whole-genome sequencing libraries were prepared using the KAPA HyperPlus kit (Roche) as previously described (Gherman Uritskiy et al., 2019) and sequenced with paired 150bp reads on the HiSeq 2000 platform at the Johns Hopkins Genetic Resources Core Facility (GRCF). Total RNA-seq libraries were prepared with the

SMARTer Stranded RNA-seq kit (Takara & Bell), using 25ng of RNA input and 12 cycles for library amplification, as previously described (G. Uritskiy et al., 2019). We sequenced 21 libraries from replicate samples from 2016 and 24 libraries from replicate samples from 2017 (**Table S1**).

Metagenomic sequence processing and MAG (metagenome-assembled genome) recovery. The de-multiplexed shotgun metagenomic sequencing reads were processed with the complete metaWRAP v0.8.2 pipeline (Uritskiy, DiRuggiero, & Taylor, 2018) with recommended databases on a UNIX cluster with 48 cores and 1024GB of RAM available. This study used the publicly available metagenomic assembly, annotation, and metagenome-assembled genomes (MAGs) from previous work (Gherman Uritskiy et al., 2019). MAGs with minimum completion of 70% and maximum contamination of 5%, as determined with CheckM (Parks, Imelfort, Skennerton, Hugenholtz, & Tyson, 2015), were used in this study. Detailed scripts for the entire analysis pipeline can be found at https://github.com/ursky/timeline_paper.

SnapT for sRNA community identification. An analytic pipeline, SnapT for Small ncRNA Annotation Pipeline for (meta)Transcriptomic data, was adapted from our previous work (D. R. Gelsinger & J. DiRuggiero, 2018) to find, annotate, and quantify intergenic and antisense sRNA transcripts from transcriptomic or metatranscriptomic data. In brief, de-novo transcripts were assembled from RNA reads mapped to the metagenomic assembly, and transcripts that could not be explained by any protein-coding region and did not encode for peptides were extracted and further validated as sRNAs. Detailed scripts for the pipeline can be found at https://github.com/ursky/SnapT and search criteria were as follows: intergenic transcripts were at least 30 nt away from any gene or ORF on both

strands; antisense transcripts were 30 nt away from any gene on their strand, but overlapped with a gene on the opposite strand by at least 10 nt; small peptides (<100 nt) were not counted as genes if they were encoded in a transcript that was more than 3 times their length; non-coding transcripts could not contain any reading frame greater than 1/3 of their lengths; predicted non-coding transcripts near contig edges were discarded and the minimum distance to the edge of a contig was dynamically computed such that the tips of contigs were not statistically enriched in annotated ncRNAs; small ncRNAs were between 50 nt and 500 nt in length; sRNA transcripts could not have significant homology with any protein in the NCBI_nr database (query cover>30%, Bitscore>50, evalue<0.0001, and identity>30%) and with any tRNA, RNase P, or signal recognition particle (SRP) model in the Rfam non-coding RNA database.

Taxonomic assignment and distribution of sRNAs. The taxonomic origin of each annotated sRNA was taken to be as that of the contig on which it lay. The taxonomy of each contig was estimated by taking the weighted average of the taxonomic assignment of the genes encoded on it, as determined through the JGI IMG functional and taxonomic annotation service (https://img.jgi.doe.gov/).

Metatranscriptomic Correlation and Differential Expression Analysis. We used a read count-based differential expression analysis to identify differentially expressed sRNA and mRNA transcripts. The program featureCounts (Liao, Smyth, & Shi, 2014) was used to rapidly count reads that map to the assembled RNA transcripts (described above) as previously described (D. R. Gelsinger & J. DiRuggiero, 2018). In order to account for organism abundance changes (as opposed to true transcript changes), we normalized the transcript read counts to the total read counts from the contig on which the transcript lies

on. The read counts were then used in the R differential expression software package DESeq2 (Love, Huber, & Anders, 2014) to calculate differential expression by determining the difference in read counts between 2016 normalized read counts from 2017 normalized read counts. The differentially expressed RNAs were filtered based on the statistical parameter of False Discovery Rate (FDR) and those that were equal to or under an FDR of 5% were classified as true differentially expressed transcripts. We carried out differential expression analysis using a pairwise Wald test to find any possible differences between years (Love et al., 2014). In parallel, normalized expression values were calculated using stringtie in transcripts per million (TPM). TPM of transcripts was normalized in the same way as read counts, except using contig TPM. TPM of transcripts was used for ranking of expression within samples as opposed to differential expression analysis.

Regulatory element motif identification of sRNAs, structure and target prediction. 50 nucleotides upstream from the sRNA transcript start coordinates were searched for transcription motifs (BRE and TATA-box for archaea and -35 and -10 consensus sequences for bacteria) using both multiple sequence alignments and visualization with WebLogo and motif searching with MEME (D. R. Gelsinger & J. DiRuggiero, 2018). Conserved sRNAs were identified using blastn against the NCBI nt database. Secondary structures of conserved sRNAs were predicted using sRNAs that had an e-value maximum of 1E-3, sequence similarity of 70% or more, and 50% or more coverage with a NCBI nt database blastn hit; a minimum of 14 alignments was used in the program LocARNA using global alignment settings (Will, Joshi, Hofacker, Stadler, & Backofen, 2012). Lastly, putative targets were predicted for itsRNAs by searching for optimal sRNA-mRNA hybridization using the IntaRNA program with the no seed parameter (Mann, Wright, & Backofen, 2017)

and the reference genes for each respective MAG. Targets were ranked by lowest p-value. Expression levels for putative targets of antisense sRNAs were obtained from co-expression analysis of transcripts (D. R. Gelsinger & J. DiRuggiero, 2018). The sRNA and putative target mRNA TPM expression values were tracked across the replicates, and the Pearson correlation was computed.

Enrichment cultures. Three types of culture medium were inoculated in triplicate with ~2 g of grounded halite colonization zones and incubated at 42°C with shaking at 220 rpm (Amerex Gyromax 737) for 1 to 2 weeks. Cells were harvested by centrifugation and nucleic acids extracted as described above. Media were: GN101 medium (Kish et al., 2009) containing 250 g of salt per L and 10 g of peptone as carbon source; Hv-YPC medium (Dyall-Smith, 2009) containing 250 g of salt per L and 8.5 g of yeast extract, 1.7 g of peptone, and 1.7 of casamino acids as carbon sources; and IO containing 250 g of salt and the same carbon sources as the Hv-YPC medium. The taxonomic distribution of the cultures was obtained with 16S rRNA gene sequencing as previously described (Gherman Uritskiy et al., 2019).

sRNA validation. Total RNA extracted from environmental samples and enrichment cultures was converted into cDNA using the SuperScript III First-Strand Synthesis System (ThermoFisher) using 5ng of input RNA. The cDNA was then amplified using 515F/926R 16S rRNA primers as previously described (G. Uritskiy et al., 2019). Amplicons were sequenced using Sanger sequencing (GENEWIZ).

Data availability. Raw sequencing data are available from the National Centre for Biotechnology Information under NCBI project ID PRJNA484015. The metagenome co-assembly and functional annotation are available from the JGI Genome Portal under IMG

taxon OID 3300027982. Metatranscriptome data have been deposited in NCBI's Gene Expression Omnibus and are accessible through GEO Series accession number GSE137164 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE137164); Scripts for functional annotation, statistical analyses, differential expression, and figures are available at https://github.com/ursky/srna_metatranscriptome_paper.

CONTRIBUTIONS

DRG: Conceptualization, Investigation, Field work, Methodology, Project administration, Writing – original draft, Writing – review & editing

GU: Investigation, Field work, Methodology, Project administration, Writing – original draft, Writing – review & editing

RR: Investigation – experimental validation, RT-PCR.

AM: Investigation – experimental validation, RT-PCR.

KF: Investigation – experimental validation, RT-PCR.

JDR: Conceptualization, Field work, Funding acquisition, Project administration, Supervision, Validation, Writing – review & editing

ACKNOWLEDGEMENTS

This work was supported by NASA grants 18-EXO18-0091 and NNX15AP18. We thank Alfonso Davila for support in the field.

SUPPLEMENTAL FIGURES

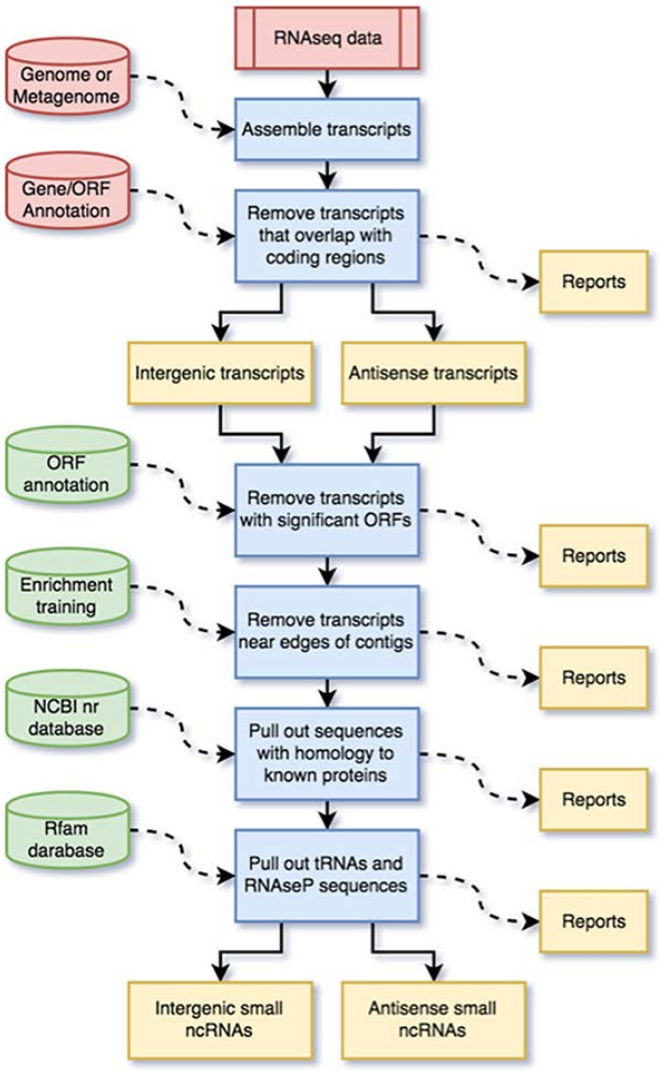

Figure 5-S1: Flow chart for SnapT methodology.

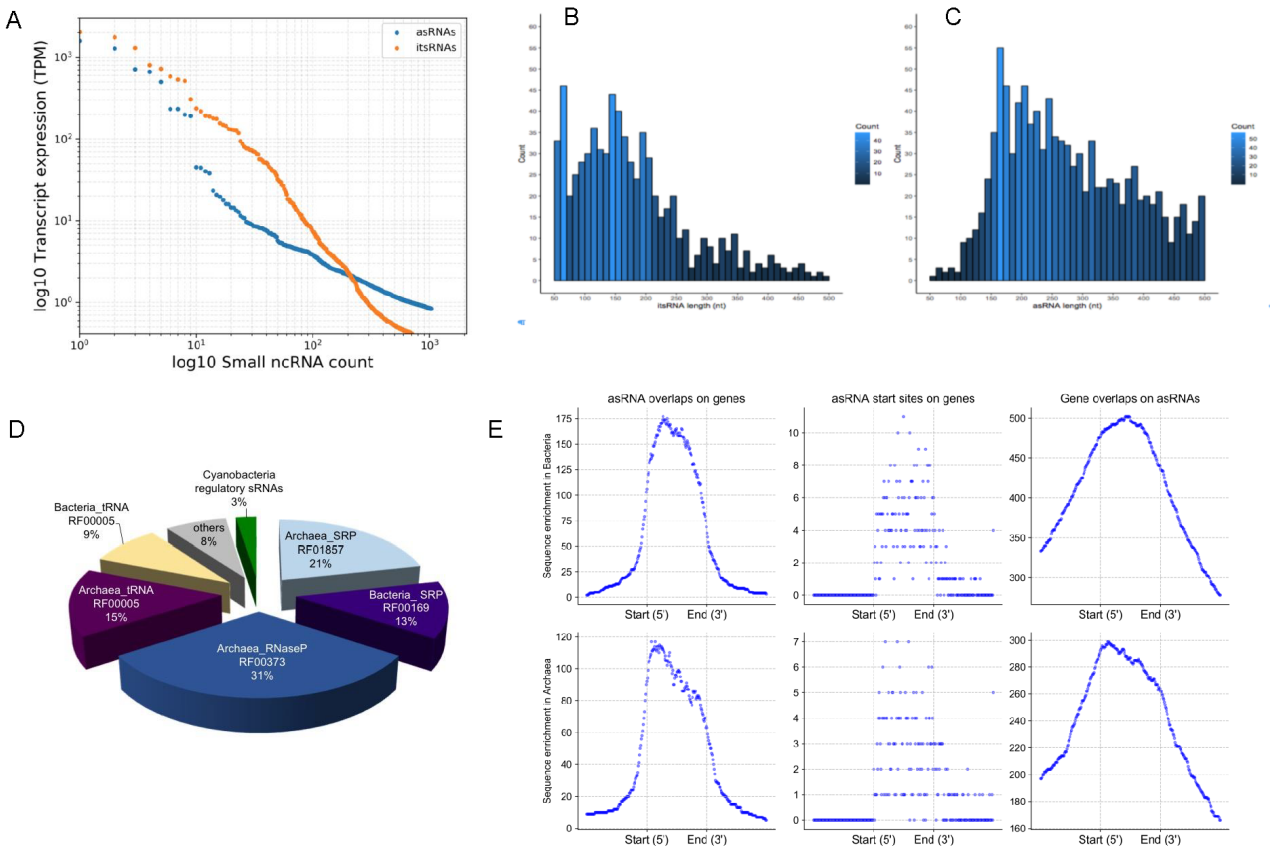

Figure 5-S2: Properties of sRNAs. (**A**) Thresholding for its and asRNAs. Ranked total expression (transcripts per million) of annotated antisense and intergenic small ncRNAs. The figure shows the linear relationship between the sRNAs expression in TPM and the number of sRNAs and how it decays. A threshold at 5x and 10x coverage was applied to its and asRNAs, respectively. Length distribution of (**B**) itsRNAs and (**C**) asRNAs. (**D**) Rfam-conserved sRNAs identified the halite community. (**E**) asRNA overlap distribution with their putative mRNAs.

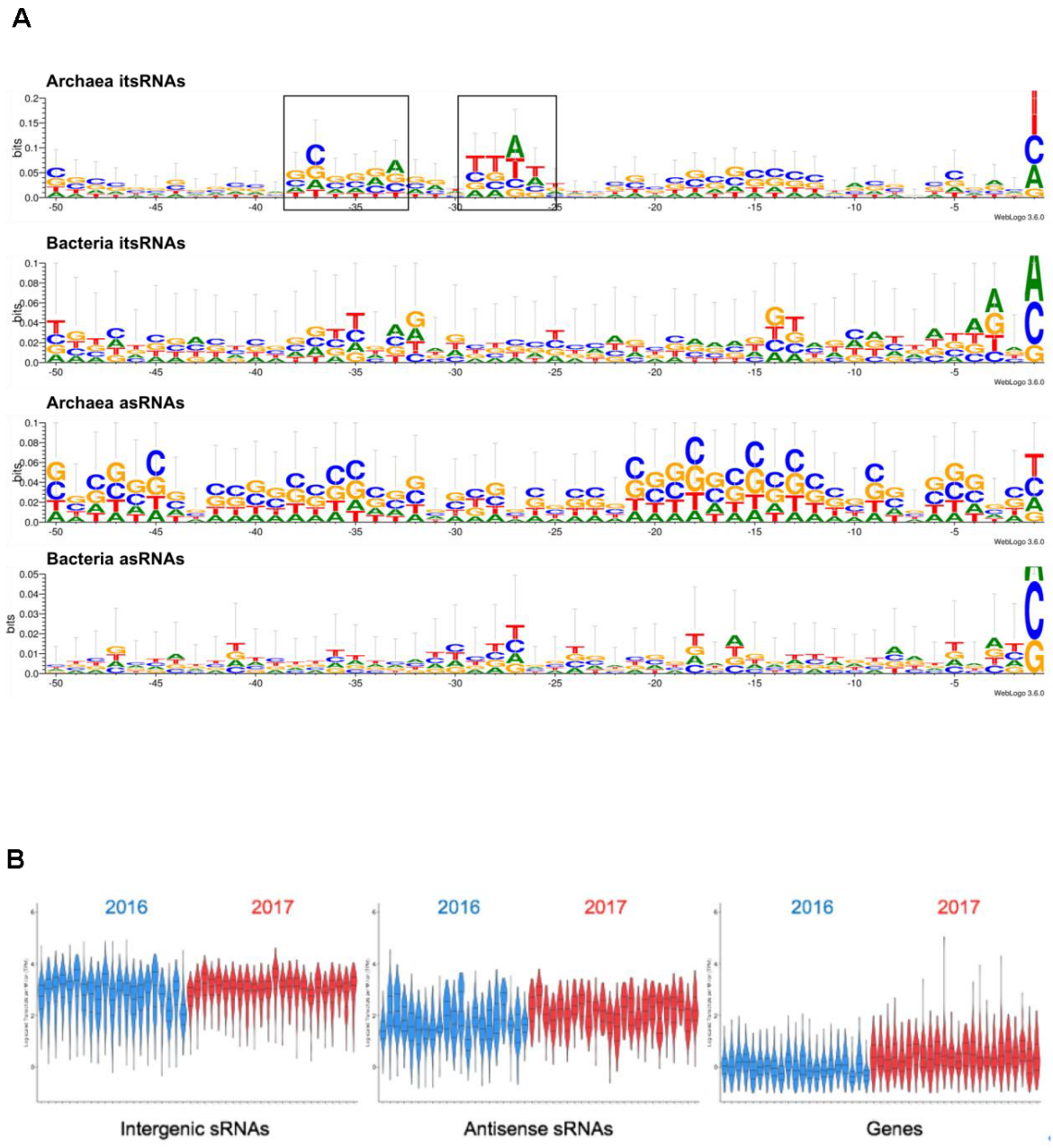

Figure 5-S3: (**A**) Regulatory regions for asRNAs and itsRNAs and for archaea and bacteria and (**B**) Expression level of itsRNAs, asRNAs, and protein encoding genes normalized by contig abundance.

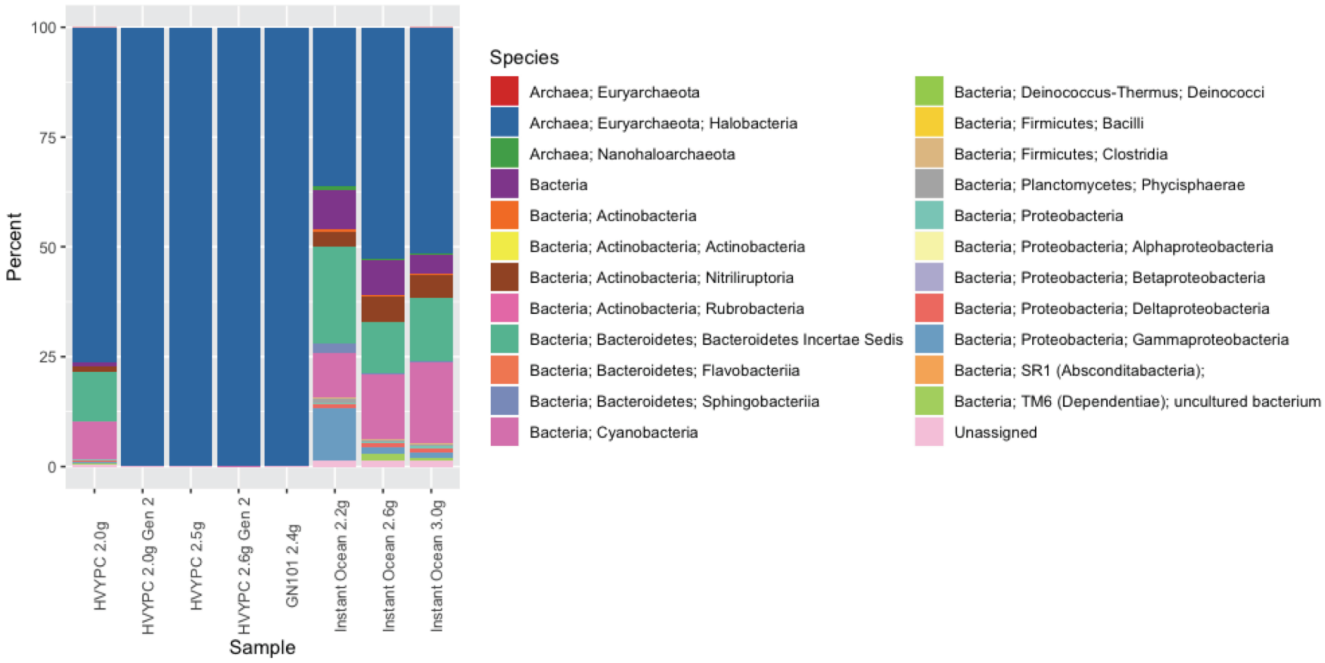

Figure 5-S4: Taxonomy distribution of halite enrichment cultures showing the relative abundance (%) of bacterial and archaeal taxa in function of the culture media.

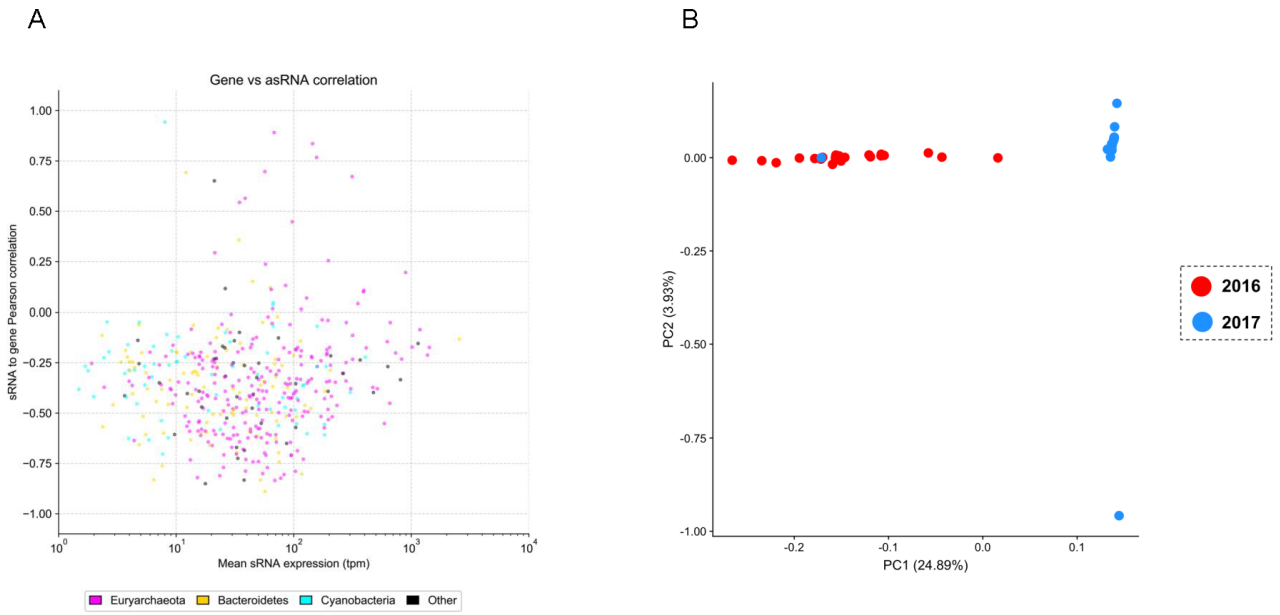

Figure 5-S5: PCA plots of (**A**) Pearson correlations in expression level between asRNAs and their putative mRNA targets across all the replicates, color coded by taxa and (**B**) expression levels of annotated genes from the metatranscriptome for all samples

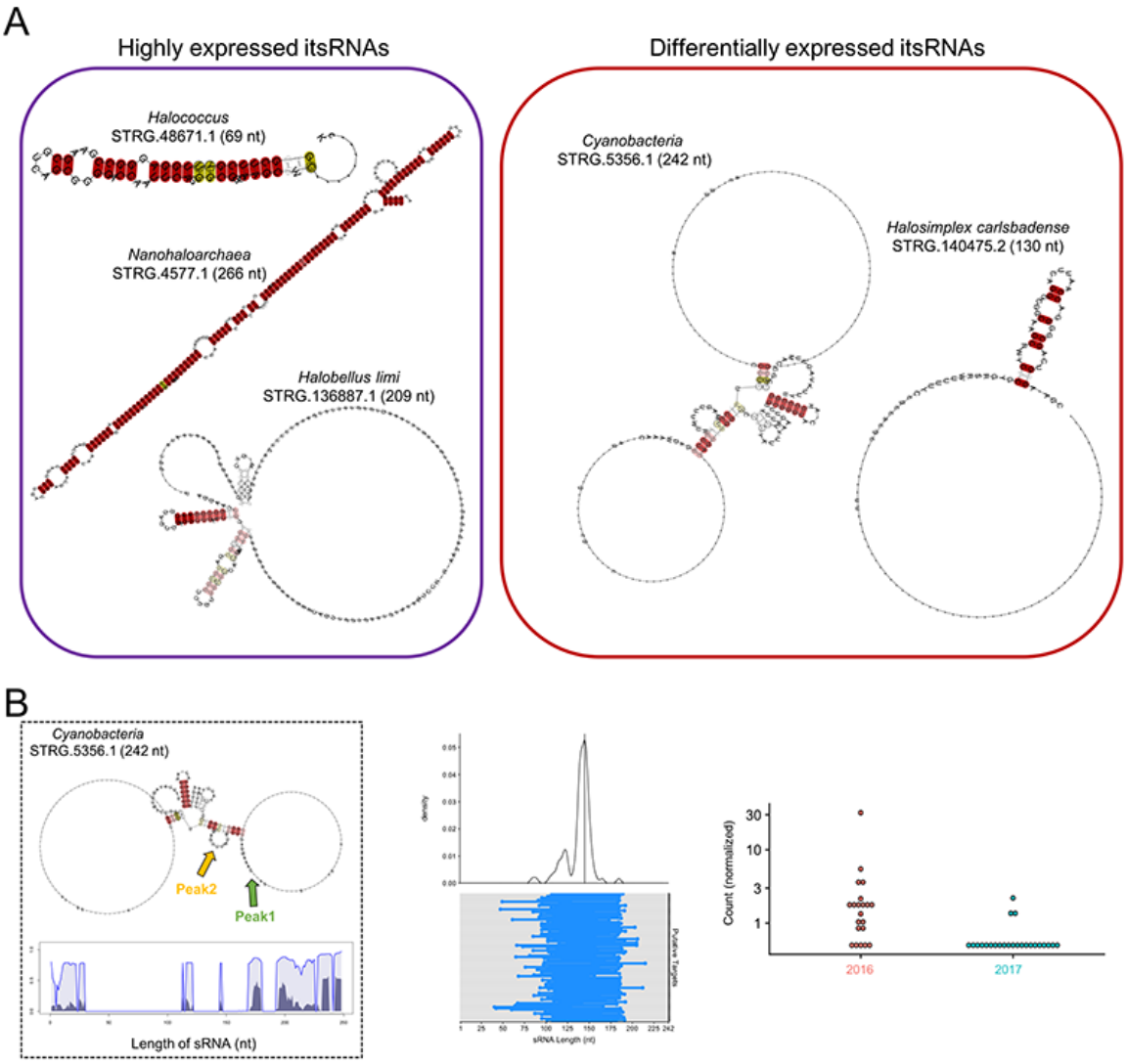

Figure 5-S6: (**A**) Additional structure prediction for highly expressed and differential. expressed itsRNAs; (**B**) Predicted structure, target identification, and expression levels for differentially expressed itsRNA STRG.5356.1 from a cyanobacterium. Structures are using minimum free energies (MFE).

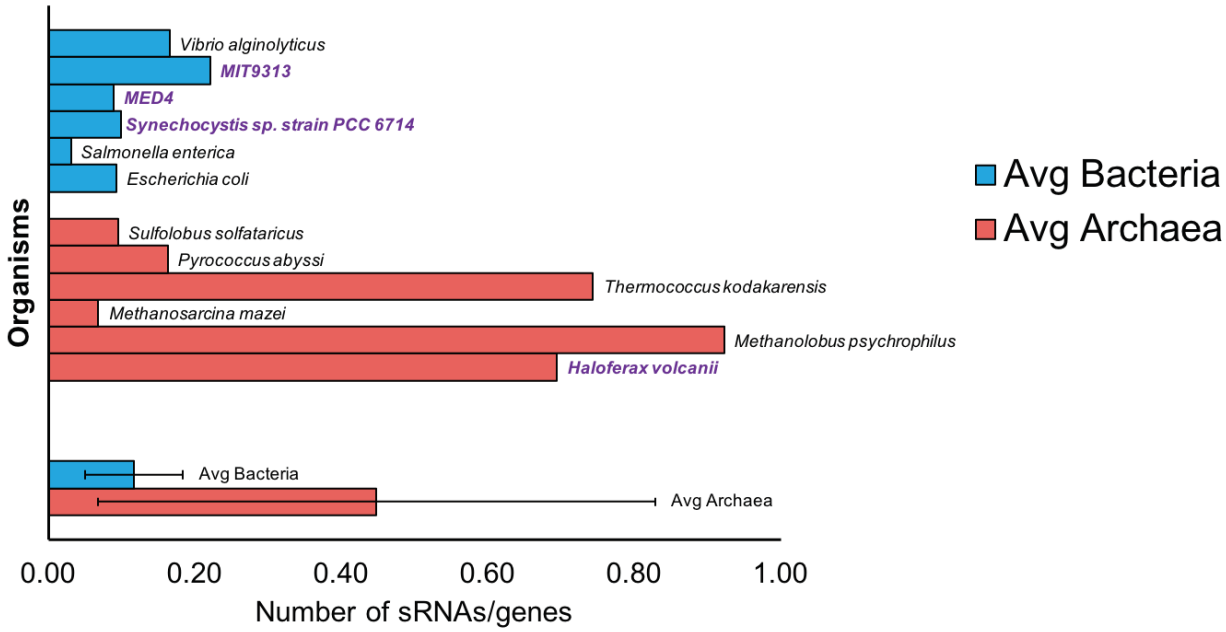

Figure 5-S7: sRNA distribution per genome of model organisms from previous studies from the literature.

Conclusions

The focus of this thesis was the elucidation of sRNAs that govern the resistance to oxidative stress in the haloarchaeon, *Haloferax volcanii*. By combining an experimental approach with computational analysis, I discovered thousands of sRNAs (both intergenic and antisense) in the transcriptome of *H. volcanii*, with hundreds specifically regulated by oxidative stress (Chapter 2). I further elucidated the regulatory mechanism of sRNAs in *H. volcanii* by focusing on an intergenic sRNA (SHOxi), which was the most up-regulated sRNA under oxidative stress. I found that SHOxi interacts, via RNA-RNA hybridization, with the malic enzyme mRNA to destabilize this abundant transcript (Chapter 4). The sRNA-mediated post-transcriptional regulation of malic enzyme under oxidative stress resulted in the reduction of NADH levels and, as a consequence, increased cell survival by reducing reactive oxygen species production in *H. volcanii*. In addition to SHOxi, I have successfully constructed knockout deletions for 3 more intergenic sRNAs that are up-regulated during oxidative stress. A similar approach to that of SHOxi is being administered to these sRNAs to uncover novel regulatory processes in the oxidative stress response of *H. volcanii*.

To determine whether SHOxi might regulate mRNA targets at the translational level, I elucidated the translational landscape in *H. volcanii* by developing, for the first time in Archaea, the high throughput technique ribosome profiling (Ribo-seq) (Chapter 3). I used Ribo-seq to verify the coding-capacity (or lack thereof) of the sRNAs discovered in Chapter 2 and found that a minority of sRNAs were associated with

ribosomes but not translated. Next, I coupled Ribo-seq with translation inhibitors to reveal essential features of translation in *H. volcanii*. This included the ribosome footprint size (27 nt), which is more similar to that of Eukaryotes, different ribosome states in translation in a leaderless transcriptome, and ribosome pausing; all of which are likely to play important regulatory roles in translation regulation during oxidative stress. Moreover, Ribo-seq revealed novel translation start sites (aTSS) with a subset of these aTSS being regulated at the translational level during oxidative stress. Finally, using the methods developed for *H. volcanii*, I explored the hypothesis that sRNAs play essential roles in the natural environment. By developing a computational pipeline, SnapT, I used metatranscriptomic data from halite rocks, which are abundant in haloarchaea, to identified numerous sRNAs from diverse microorganisms in the community with potentially important regulatory roles (Chapter 5). My thesis research has greatly advanced our understanding of the mechanism by which an archaeal sRNA can regulate the oxidative stress response. Moreover, by developing ribosome profiling to elucidate the translational landscape of an archaeon, I have paved the way for studying translation regulation, for the first time, in the third domain of life. I will now review the implications of these findings and propose future directions for this research.

Large-scale identification of sRNAs targets

To date, very few of the newly reported candidate sRNAs in the Archaea have been functionally characterized (Buddeweg et al., 2017; Jager et al., 2012; Daniela Prasse et al., 2017) and many questions remain: *what are the targets of the multitude of sRNAs discovered in the Archaea? What are the regulatory effects of these sRNAs? And more*

importantly, what type of molecular mechanisms can we expect in a domain of life where information processing systems are a mosaic of bacterial and eukaryal systems (Bell & Jackson, 1998)?

Target identification of sRNAs, especially intergenic RNAs, is a difficult task due to partial base pairing with multiple targets by a single sRNA. In this thesis, I provided a strategy to identify both sRNAs and the targets of those sRNAs using reverse molecular genetics. Specifically, I validated and characterized one target of the sRNA SHOxi using a combination of RNA-seq and Ribo-seq. Elucidating the post-transcriptional regulation of malic enzyme mRNA by SHOxi provided evidence that sRNAs are crucial for the survival of an archaeon to stress and that fine-tuning of metabolism is a core strategy to mitigate damage from oxidative stress. This is in agreement with the observation in Chapter 2 where we found an enrichment of metabolic genes that are down-regulated transcriptionally during oxidative stress. To further understand the effects of SHOxi more potential targets need to be characterized.

Beside malic enzyme mRNA, I identified other potential SHOxi targets in the same datasets. These other putative targets include several transcription factors with unknown functions, which suggests that SHOxi may be a master regulator with large downstream effects in gene regulatory networks. Interestingly, all of these transcription factor mRNAs have increased transcript levels in the presence of SHOxi, suggesting that they might be stabilized by SHOxi rather than degraded like malic enzyme mRNA. A dual functioning sRNA has been observed in other Archaea (Jager et al., 2012; Prasse et al., 2017), suggesting that this might also be the case for SHOxi in *H. volcanii*. In contrast, several other mRNAs were down-regulated in the presence of SHOxi, including a sugar ABC

transporter operon. Elucidating whether these mRNAs are true targets of SHOxi and the regulatory mechanism of an operon promises to be interesting because it is thought that *H. volcanii* does not use Shine-Dalgarno, a regulatory feature commonly used on operons in other Archaea (Kramer et al., 2014). A similar strategy to the one used in Chapter 4 would potentially elucidate such regulatory mechanisms, but it will only be one target at a time and require extensive molecular analysis. In contrast, the strategies described below provide a high throughput option to elucidate all targets of SHOxi during oxidative stress and to screen for other sRNAs identified in Chapter 2.

New high-throughput strategies have recently been developed in Bacteria and Eukarya to directly identify RNA-RNA duplexes *in vivo* using MS2 hairpins (MAPS) (Lalaouna et al., 2017) or psoralen-mediated crosslinking (PARIS/LIGR/SPLASH). One of these methods, Sequencing of Psoralen crosslinked Ligated, and Selected Hybrid (SPLASH) (Aw *et al.*, 2016) involves five essential steps: (1) specific *in vivo* cross-linking of RNA-RNA duplexes using psoralen (or derivatives), (2) enrichment of duplexes using biotin-streptavidin methods and degradation of non-duplexed RNA, (3) ligation of an adapter loop to form a chimeric RNA molecule, (4) reverse crosslinking, and (5) high-throughput sequencing of the chimeric RNAs. In contrast to alternative methods for RNA-RNA interactions, such as CLASH and CLIP (Weidmann et al., 2016), SPLASH does not require an sRNA-interacting protein in complex with the RNA duplexes, a component that is still unresolved in the Archaea. Research from these approaches can potentially identify entire sRNA regulons and provide information on sRNA-mRNA "seed" regions and structural binding motifs that can be used to build archaeal-specific sRNA target prediction models. Potential problems to these methods are that psoralen preferential cross-linking might

result in missing targets, false positive, and/or masking of lowly expressed sRNAs by highly abundant RNAs and their interactions. Despite these drawbacks, these methods are a step forward and they provide useful tools for sRNA biology. However, it is important to note that additional experiments will be necessary to validate these experimentally derived RNA-RNA interactions (i.e. mobility shift assays).

Identification of RNA-binding proteins and RNases in sRNA-mediated regulation

RNA-binding proteins (RBP) are essential for sRNA-mediated regulation in the other domains of life; thus, it is likely that such proteins also play a role in the Archaea. Co-immuno-precipitation with the Lsm protein, an archaeal homolog of the bacterial sRNA chaperone Hfq, was used to "capture" sRNAs *in vitro* (Fischer et al., 2010). These experiments were not replicated and the role of Lsm – or any other RNA-binding protein - remains to be elucidated in the Archaea. I was unable to obtain a deletion mutant of Lsm using classical reverse molecular genetics, suggesting its necessary function in the cell. Recently, a CRISPR interference system was developed in *H. volcanii* (Stachler et al., 2016), which could be used to knockdown Lsm expression instead and assess whether Lsm is required for sRNAs interactions with their targets. Homologs of eukaryotic sRNA interacting proteins (Argonautes) have been found in the Archaea, but rather than RNA interference, a defensive role against foreign genetic elements have been proposed (Willkomm et al., 2017). An alternative strategy for identification of RNA-binding proteins involved in sRNA regulation is to use a *bona fide* sRNA as a bait to pull-down interacting proteins. With the elucidated role of SHOxi and its importance in stress survival, it is a top candidate to carry out such an experiment. In brief, SHOxi would be transcribed *in vitro*

and covalently bonded to biotin. Next, whole lysate from *H. volcanii* would be flowed onto the immobilized SHOxi, and the bound proteins will be identified by mass spectrometry. Future work in the laboratory will continue in this direction using SHOxi, and other sRNAs that I have discovered, as protein baits.

Ribonuclease degradation of mRNAs mediated by sRNAs (both intergenic and antisense) is a hallmark of bacterial sRNA-mRNA regulation. Ribonucleases have been found to play large-scale roles in 5'-3' directed mRNA decay in the Archaea, including enzymes such as CPRSF2 in *S. acidocaldarius* and RNase J in *M. jannaschii* (Levy et al., 2011), raising the question whether there is an intersection between these RNases and sRNA regulation. Some Archaea have very short or no 5' UTRs on mRNAs, such as *H. volcanii*, suggesting that the 5' UTR in these Archaea do not carry information regarding translation initiation or transcript degradation. The observation that the majority of sRNAs in *H. volcanii* interact within the CDS of targets (Gelsinger & DiRuggiero, 2018) could indicate that, rather than an RNase with exoribonucleolytic activity (CPRSF2/RNase J), an RNase with endoribonucleolytic activity, such as CPSF1 in *M. jannaschii*, could be the major RNase interacting with sRNA-mRNA duplexes (Levy et al., 2011). Despite the identification of several RNases in Archaea, many of their targets remain to be elucidated and there is still no evidence that these enzymes interact or bind sRNAs. Therefore, questions about what ribonucleoprotein complexes are involved in archaeal sRNA regulation and their mechanistic roles remain unanswered. The finding that the sRNA SHOxi destabilizes malic enzyme mRNA during oxidative stress in *H. volcanii* strongly suggests that an RNase is involved in the post-transcriptional regulation of this transcript. Therefore, SHOxi sRNA provides an excellent model for investigating the hypothesis that

RNases are involved in sRNA-mediated regulation in Archaea. The strategy previously described for RBP identification via a biotin tethered SHOxi might also reveal the identity of RNases involved.

Transcription regulation by sRNAs in Archaea

I found that the majority of the sRNAs in *H. volcanii* were antisense to coding genes, and intriguingly there was a strong negative correlation between up-regulated antisense sRNAs (asRNAs) and down-regulated transposons encoded on the opposite strand during oxidative stress. This suggests a potential negative regulatory effect by the asRNAs onto transposon mRNAs. Limited work has been done on asRNAs in all domains of life, but this negative correlation is similar to transposon regulatory processes observed in Bacteria (Ellis et al., 2015). A sophisticated genetic system is necessary to uncover the regulatory effects of these asRNAs without perturbing the potential target gene. A common strategy to study asRNAs in bacteria is a dual plasmid system to express the asRNA on its own plasmid and the target mRNA (i.e. transposons) on a second plasmid. In order to screen with high throughput, a reporter (i.e. GFP) could be C-terminally linked to the gene and regulation can be measured by a fluorescence output. Currently, a dual plasmid system in *H. volcanii* does not exist; I initiated the development of a dual plasmid system in *H. volcanii* that could be used, in future work, to validate whether asRNAs regulate transposons during oxidative stress. Studying this mechanism is particularly intriguing because it has been shown that transposons are opportunistic during stress conditions and wreak havoc by hopping around in the genome causing double strand breaks; hence a need to be silenced (Capy et al., 2000; Wheeler, 2013; Whitehead et al., 2006). A functional

enrichment of IS4 transposon genes (30% of all transposons) being down-regulated suggests that transposon activity is tightly regulated during oxidative stress in *H. volcanii*. Taken together, I hypothesize that this could be a potential mechanism for oxidative stress resistance.

Translation regulation by sRNAs in Archaea

In Chapter 3, by developing ribosome profiling and elucidating the translational landscape in *H. volcanii,* I provided a tool for the community to study translation regulation in Archaea, a vastly understudied field of research. Regulation by masking the ribosome binding site on mRNAs has been proposed in methanogens (Prasse et al., 2017; Buddeweg et al., 2018). More recently, a sRNA in *H. volcanii* (s194) was proposed to regulate the translation of an mRNA through an alternative mechanism by directly binding the ribosome (Wyss et al., 2018). This all indicates that archaeal sRNAs could regulate translation. Combining Ribo-seq with SHOxi deletion mutants with mRNA-seq, I was able to calculate translation efficiency (TE) for all mRNAs in *H. volcanii.* I found that there were significant TE decreases of mRNAs in the presence of SHOxi, suggesting that this sRNA can potentially regulate transcripts at the level of translation. This work is now the foundation for the thesis of a graduate student in the laboratory, who is validating these observations by GFP tagging the putative targets and measuring fluorescence, as a proxy for translation, under SHOxi overexpression in *H. volcanii*.

I used ribosome profiling, combined with HHT-treatment, to distinguish between transcripts that were being translated versus those that were not in order to validate sRNAs I identified in Chapter 2. If ribosome density was altered towards a TSS with HHT-

treatment, then it was categorized as a transcript being translated. I found that a subset of these sRNAs, including s194 and SHOxi, exhibited ribosome density along their transcript. Of these, I found 3 sRNAs potentially encoding small proteins while the others were not (i.e. SHOxi, s194). Unraveling why ribosomes are associating with sRNAs, but not translating them, could provide a new avenue for research on archaeal sRNA regulation. Future experiments to elucidate this puzzling observation could include knocking out these sRNAs and monitoring translation differences with Ribo-seq.

Alternative reading frame usage as a means of translation regulation

Using ribosome profiling in conjunction with translation inhibitors, I also identified potential protein-coding sequences in contexts that are usually ignored in other datasets, such as TSS that are extremely small, or those that are within or antisense to known ORFs. These alternative proteins have also been found in both bacterial and eukaryotic contexts using ribosome profiling (Ingolia et al., 2011; Meydan et al., 2019; Weaver et al., 2019). In particular, we found hundreds of alternative TSS such as: initiation sites up-stream of annotated TSS (N-terminal extensions), initiation sites within annotated ORFs that could produce truncated proteins (internal in-frame) or completely new proteins (internal out-of-frame), and small ORFs (<50 amino acids). These alternative proteins may have been previously obscured in proteomic data but may play important roles in cell physiology and stress response. N-terminally extended and internal initiation could regulate their corresponding ORFs by altering translation efficiency of the annotated protein, as proposed in Bacteria and Eukarya (Ingolia et al., 2011; Meydan et al., 2019). The differential translation regulation of some of these aTSS during oxidative stress validated their

existence in the cell and suggested that they may be necessary for cells to adapt to their environment. It is also possible that not all identified aTSS produce a protein and that they may be remnants of problems in translation within the cell, such as inaccurate start site picking regulated by Inr elements, thus verification of protein products is necessary.

The verification of these putative alternative proteins is more challenging in Archaea, compared to Bacteria and Eukarya, due to less developed genetic tools and biochemical restrictions due to proteomic adaptations to extreme environments. Moreover, the identification of these alternative proteins is not straightforward using mass spectrometry, in particular, for the internal in-frame where it is difficult to distinguish signal from the annotated ORF and for smORFs where there is limited chance for trypsin digestion. Even so, the N-terminal extensions, the internal out-of-frame, and potentially some smORFs could be searched in mass spectrometry data depending on preparation and trypsin digestion. Future work in the laboratory will aim to verify these aTSS products using such an approach or by using an alternative approach, first developed in Bacteria, which uses Gibson cloning to FLAG-tag these potential alternative proteins and verify their existence through Western blotting (Weaver et al., 2019).

Future prospects of sRNA research in Archaea

It has become clear that large-scale gene regulation is essential for Archaea to thrive in their environments and that sRNAs play key roles in those regulatory processes. Recent work on the molecular biology of sRNAs in archaea, including this thesis, have opened new areas of research and generated many more questions. Some of the more outstanding/pressing questions include the role of more than 1100 cis-sense sRNAs

recently discovered in *H. volcanii* (Babski et al., 2016), the prevalence of regulatory tRNA-derived fragments in the archaea (Wyss et al., 2017; Heyer et al., 2012) and the potential for sRNAs to encode small peptides such as in bacteria and eukarya (Bronsard et al., 2017). Moreover, *in vivo* quantitative measurements of sRNA-mediated regulation, such as those currently being made in bacteria (Bobrovskyy et al., 2017), and now recently in archaea (Kliemt et al., 2019; Wyss et al., 2018), are necessary to understand, at a system-level, how sRNA-based regulation is integrated within a cell's regulatory networks. Finally, assessing to what degree sRNAs play in gene regulatory networks within a microbial community, including but not limited to influencing community composition and inter-cellular communication, will undoubtedly increase our understanding of microorganisms in their natural environments.

Bibliography

REFERENCES

Ajay Sharma, Elena K. Gaidamakova, Olga Grichenko, Vera Y. Matrosova, Veronika Hoeke, Polina Klimenkova, Isabel H. Conze, Robert P. Volpe, Rok Tkavcb, Cene Gostinþar, Nina Gunde-Cimerman, Jocelyne DiRuggiero, Igor Shuryakg, Andrew Ozarowskih, Brian M. Hoffmana, Michael J. Daly. "Across the Tree of Life, Radiation Resistance Is Governed by Antioxidant Mn2+, Gauged by Paramagnetic Resonance." *Submitted* (2017). Print.

Albers, Sonja-Verena, and Ken F. Jarrell. "The Archaellum: How Archaea Swim." *Frontiers in Microbiology* 6 (2015): 23. Print.

Allers, T, Barak, S., Liddell, S., Wardell, K., & Mevarech, M. (2010). Improved strains and plasmid vectors for conditional overexpression of His-tagged proteins in Haloferax volcanii. *Appl Environ Microbiol*, *76*(6), 1759–1769. https://doi.org/AEM.02670-09 [pii]10.1128/AEM.02670-09

Allers, T, & Mevarech, M. (2005). Archaeal genetics - the third way. *Nat Rev Genet*, *6*(1), 58-73.

Allers, Thorsten, Ngo, H.-P., Mevarech, M., & Lloyd, R. G. (2004). Development of Additional Selectable Markers for the Halophilic Archaeon Haloferax volcanii Based on the leuB and trpA Genes. *Appl. Environ. Microbiol.*, *70*(2), 943–953. Retrieved from http://aem.asm.org/cgi/content/abstract/70/2/943

Altschul, Stephen F., et al. "Basic Local Alignment Search Tool." *Journal of Molecular Biology* 215.3 (1990): 403-10. Print.

Altuvia, S. "Regulatory Small Rnas: The Key to Coordinating Global Regulatory Circuits." *J Bacteriol* 186.20 (2004): 6679-80. Print.

Altuvia, S., Weinstein-Fischer, D., Zhang, A., Postow, L., & Storz, G. (1997). A small, stable RNA induced by oxidative stress: role as a pleiotropic regulator and antimutator. *Cell*, *90*(1), 43–53. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9230301

Aravalli, R. N., She, Q. X., & Garrett, R. A. (1998). Archaea and the new age of microorganisms. *Trends in Ecology & Evolution*, *13*, 190–194.

Aw, J. G. A., Shen, Y., Wilm, A., Sun, M., Lim, X. N., Boon, K.-L., … Wan, Y. (2016). In Vivo Mapping of Eukaryotic RNA Interactomes Reveals Principles of Higher-Order Organization and Regulation. *Molecular Cell*, *62*(4), 603–617. https://doi.org/10.1016/j.molcel.2016.04.028

Babski, J, Tjaden, B., Voss, B., Jellen-Ritter, A., Marchfelder, A., Hess, W. R., & Soppa, J. (2011). Bioinformatic prediction and experimental verification of sRNAs in the haloarchaeon Haloferax volcanii. *RNA Biol*, *8*(5), 806–816. https://doi.org/10.4161/rna.8.5.16039

Babski, Julia, Haas, K. A., Näther-Schindler, D., Pfeiffer, F., Förstner, K. U., Hammelmann, M., Soppa, J. (2016). Genome-wide identification of transcriptional start sites in the haloarchaeon Haloferax volcanii based on differential RNA-Seq (dRNA-Seq). *BMC Genomics*, *17*(1), 629. https://doi.org/10.1186/s12864-016-2920-y

Baek, Y. M., Jang, K.-J., Lee, H., Yoon, S., Baek, A., Lee, K., & Kim, D.-E. (2019). The bacterial endoribonuclease RNase E can cleave RNA in the absence of the RNA

chaperone Hfq. *The Journal of Biological Chemistry*, *294*(44), 16465–16478. https://doi.org/10.1074/jbc.RA119.010105

Bailey, T. L., et al. "Meme Suite: Tools for Motif Discovery and Searching." *Nucleic Acids Res* 37.Web Server issue (2009): W202-8. Print.

Balasubramanian, D., Ragunathan, P. T., Fei, J., & Vanderpool, C. K. (2016). A Prophage-Encoded Small RNA Controls Metabolism and Cell Division in <em>Escherichia coli</em&gt. *MSystems*, *1*(1). Retrieved from http://msystems.asm.org/content/1/1/e00021-15.abstract

Baliga, N. S., Bjork, S. J., Bonneau, R., Pan, M., Iloanusi, C., Kottemann, M. C., … DiRuggiero, J. (2004). Systems level insights into the stress response to UV radiation in the halophilic archaeon Halobacterium NRC-1. *Genome Res*, *14*(6), 1025–1035. https://doi.org/10.1101/gr.19935041993504

Bandyra, K. J., Said, N., Pfeiffer, V., Górna, M. W., Vogel, J., & Luisi, B. F. (2012). The seed region of a small RNA drives the controlled destruction of the target mRNA by the endoribonuclease RNase E. *Molecular Cell*, *47*(6), 943–953. https://doi.org/10.1016/j.molcel.2012.07.015

Barshishat, S., Elgrably-Weiss, M., Edelstein, J., Georg, J., Govindarajan, S., Haviv, M., … Altuvia, S. (2018). OxyS small RNA induces cell cycle arrest to allow DNA damage repair. *The EMBO Journal*, *37*(3), 413–426. https://doi.org/10.15252/embj.201797651

Bao, G. H., M. J. Wang, T. G. Doak, and Y. Z. Ye. 2015. "Strand-specific community RNA-seq reveals prevalent and dynamic antisense transcription in human gut

microbiota." *Frontiers in Microbiology* 6. doi: ARTN 89610.3389/fmicb.2015.00896.

Bell, S. D., & Jackson, S. P. (1998). Transcription and translation in Archaea: a mosaic of eukaryal and bacterial features. *Trends in Microbiology*, *6*(6), 222–228. https://doi.org/https://doi.org/10.1016/S0966-842X(98)01281-5

Benelli, D., & Londei, P. (2011). Translation initiation in Archaea: conserved and domain-specific features. *Biochemical Society Transactions*, *39*(1), 89–93. https://doi.org/10.1042/BST0390089

Berlett, B. S., & Stadtman, E. R. (1997). Protein oxidation in aging, disease, and oxidative stress. *J Biol Chem*, *272*(33), 20313–20316.

Bernick, D. L., Dennis, P. P., Lui, L. M., & Lowe, T. M. (2012). Diversity of Antisense and Other Non-Coding RNAs in Archaea Revealed by Comparative Small RNA Sequencing in Four Pyrobaculum Species. *Front Microbiol*, *3*, 231. https://doi.org/10.3389/fmicb.2012.00231

Bobrovskyy, M., Frandsen, J. K., Zhang, J., Poddar, A., Azam, M. S., Henkin, T. M., … Vanderpool, C. K. (2017). Determinants of target prioritization and regulatory hierarchy for the bacterial small RNA SgrS. *BioRxiv*.

Bonneau, R., Facciotti, M. T., Reiss, D. J., Schmid, A. K., Pan, M., Kaur, A., … Baliga, N. S. (2007). A predictive model for transcriptional control of physiology in a free living cell. *Cell*, *131*(7), 1354–1365. https://doi.org/S0092-8674(07)01416-X [pii]10.1016/j.cell.2007.10.053

Brenneis, M., Hering, O., Lange, C., & Soppa, J. (2007). Experimental characterization of Cis-acting elements important for translation and transcription in halophilic

archaea. *PLoS Genet*, *3*(12), e229. https://doi.org/10.1371/journal.pgen.0030229

Bronsard, J., Pascreau, G., Sassi, M., Mauro, T., Augagneur, Y., & Felden, B. (2017). sRNA and cis-antisense sRNA identification in Staphylococcus aureus highlights an unusual sRNA gene cluster with one encoding a secreted peptide. *Scientific Reports*, *7*(1), 4565. https://doi.org/10.1038/s41598-017-04786-3

Buddeweg, A., Sharma, K., Urlaub, H., & Schmitz, R. A. (2017). sRNA41 affects ribosome binding sites within polycistronic mRNAs in Methanosarcina mazei Gö1. *Molecular Microbiology*, n/a-n/a. https://doi.org/10.1111/mmi.13900

Busch, A., Richter, A. S., & Backofen, R. (2008). IntaRNA: efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions. *Bioinformatics*, *24*(24), 2849–2856. https://doi.org/10.1093/bioinformatics/btn54

Bush, A. I. (2000). Metals and neuroscience. *Current Opinion in Chemical Biology*, *4*(2), 184–191. https://doi.org/https://doi.org/10.1016/S1367-5931(99)00073-3

Buskirk, A. R., & Green, R. (2017). Ribosome pausing, arrest and rescue in bacteria and eukaryotes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1716), 20160183. https://doi.org/10.1098/rstb.2016.0183

Cai, Qiang, Lulu Qiao, Ming Wang, Baoye He, Feng-Mao Lin, Jared Palmquist, Sienna-Da Huang, and Hailing Jin. 2018. "Plants send small RNAs in extracellular vesicles to fungal pathogen to silence virulence genes." *Science* 360 (6393):1126. doi: 10.1126/science.aar4142.

Capy, Pierre, et al. "Stress and Transposable Elements: Co-Evolution or Useful Parasites?" *Heredity* 85.2 (2000): 101-06. Print.

Carrier, Marie-Claude, David Lalaouna, and Eric Massé. 2018. "Broadening the Definition of Bacterial Small RNAs: Characteristics and Mechanisms of Action." *Annual Review of Microbiology* 72 (1):141-161. doi: 10.1146/annurev-micro-090817-062607.

Carthew, R. W., & Sontheimer, E. J. (2009). Origins and Mechanisms of miRNAs and siRNAs. *Cell*, *136*(4), 642–655. https://doi.org/10.1016/j.cell.2009.01.035

Cech, T. R., & Steitz, J. A. (2014). The Noncoding RNA Revolution: Trashing Old Rules to Forge New Ones. *Cell*, *157*(1), 77–94. https://doi.org/10.1016/j.cell.2014.03.008

Chen, X., Quinn, A. M., & Wolin, S. L. (2000). Ro ribonucleoproteins contribute to the resistance of deinococcus radiodurans to ultraviolet irradiation. *Genes Dev.*, *14*(7), 777–782.

Chen, Y., Xue, D., Sun, W., Han, J., Li, J., Gao, R., Zuo, K. (2019). sRNA OsiA Stabilizes Catalase mRNA during Oxidative Stress Response of Deincoccus radiodurans R1. *Microorganisms*, *7*(10), 422. https://doi.org/10.3390/microorganisms7100422

Clouet-d'Orval, B. (2017). *RNA Metabolism and Gene Expression in Archaea*.

Clouet-d'Orval, B., Batista, M., Bouvier, M., Quentin, Y., Fichant, G., Marchfelder, A., & Maier, L.-K. (2018). Insights into RNA-processing pathways and associated RNA-degrading enzymes in Archaea. *FEMS Microbiology Reviews*, *42*(5), 579–613. https://doi.org/10.1093/femsre/fuy016

Coker, J. A., Dassarma, P., Kumar, J., Muller, J. A., & Dassarma, S. (2007). Transcriptional profiling of the model Archaeon Halobacterium sp. NRC-1: responses to changes in salinity and temperature. *Saline Systems*, *3*, 6.

Collart, M. A., & Weiss, B. (2019). Ribosome pausing, a dangerous necessity for co-translational events. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkz763

Conesa, Ana, et al. "A Survey of Best Practices for Rna-Seq Data Analysis." *Genome Biology* 17 (2016): 13. Print.

Crits-Christoph, A., Gelsinger, D. R., Ma, B., Wierzchos, J., Ravel, J., Ascaso, C., … DiRuggiero, J. (2016). Functional analysis of the archaea, bacteria, and viruses from a halite endolithic microbial community. *Env. Microbiol.*, *18*, 2064-2077. doi: 10.1111/1462-2920.13259.

Crooks, Gavin E., et al. "Weblogo: A Sequence Logo Generator." *Genome Research* 14.6 (2004): 1188-90. Print.

Davila, A.F., B. Gomez-Silva, A. de los Rios, C. Ascaso, H. Olivares, C.P. McKay, and J. Wierzchos. 2008. "Facilitation of endolithic microbial survival in the hyperarid core of the Atacama Desert by mineral deliquescence." *J. Geophys. Res.* 113 (G01028):G01028, doi:10.1029/2007JG000561.

Davila, A.F., I. Hawes, J. Garcia, D.R. Gelsinger, J. DiRuggiero, C. Ascaso, A. Osano, and J. Wierzchos. 2015. " In situ metabolism in halite endolithic microbial communities of the hyperarid Atacama Desert." *Front Microbiol* http://dx.doi.org/10.3389/fmicb.2015.01035.

Dahl, Jan-Ulrik, Michael J. Gray, and Ursula Jakob. "Protein Quality Control under Oxidative Stress Conditions." *Journal of Molecular Biology* 427.7 (2015): 1549-63. Print.

Darnell, Cynthia L., and Amy K. Schmid. "Systems Biology Approaches to Defining Transcription Regulatory Networks in Halophilic Archaea." *Methods* 86 (2015): 102-14. Print.

Darnell, A. M., Subramaniam, A. R., & O'Shea, E. K. (2018). Translational Control through Differential Ribosome Pausing during Amino Acid Limitation in Mammalian Cells. *Molecular Cell*, *71*(2), 229-243.e11. https://doi.org/https://doi.org/10.1016/j.molcel.2018.06.041

de Almeida, João Paulo Pereira, Ricardo Z. N. Vêncio, Alan P. R. Lorenzetti, Felipe Ten Caten, José Vicente Gomes-Filho, and Tie Koide. 2019. "The Primary Antisense Transcriptome of Halobacterium salinarum NRC-1." *Genes* 10 (4):280. doi: 10.3390/genes10040280.

De Lay, Nicholas, Daniel J. Schu, and Susan Gottesman. "Bacterial Small Rna-Based Negative Regulation: Hfq and Its Accomplices." *The Journal of Biological Chemistry* 288.12 (2013): 7996-8003. Print.

Dennis, P. P. (1997). Ancient Ciphers: Translation in Archaea. *Cell*, *89*, 107–1010.

DeVeaux, Linda C., et al. "Extremely Radiation-Resistant Mutants of a Halophilic Archaeon with Increased Single-Stranded DNA-Binding Protein (Rpa) Gene Expression." *Radiation Research* 168.4 (2007): 507-14. Print.

Dyall-Smith, M. 2009. "The Halohandbook – Protocols for haloarchaeal genetics. ." *Available at* http://www.haloarchaea/. *com/resources/halohandbook/index.html*.

Eme, L., Spang, A., Lombard, J., Stairs, C. W., & Ettema, T. J. G. (2017). Archaea and the origin of eukaryotes. *Nature Reviews Microbiology*, *15*, 711.

Eichhorn, Eric, Jan R. van der Ploeg, and Thomas Leisinger. 1999. "Characterization of a Two-component Alkanesulfonate Monooxygenase from Escherichia coli." *Journal of Biological Chemistry* 274 (38):26639-26646.

Esquivel, Rianne N., and Mechthild Pohlschroder. "A Conserved Type Iv Pilin Signal Peptide H-Domain Is Critical for the Post-Translational Regulation of Flagella-Dependent Motility." *Molecular Microbiology* 93.3 (2014): 494-504. Print.

Facciotti, M. T., Pang, W. L., Lo, F. Y., Whitehead, K., Koide, T., Masumura, K., … Baliga, N. S. (2010). Large scale physiological readjustment during growth enables rapid, comprehensive and inexpensive systems analysis. *BMC Systems Biology*, *4*, 64. https://doi.org/10.1186/1752-0509-4-64

Finstad, K. M., A. J. Probst, B. C. Thomas, G. L. Andersen, C. Demergasso, A. Echeverria, R. G. Amundson, and J. F. Banfield. 2017. "Microbial Community Structure and the Persistence of Cyanobacterial Populations in Salt Crusts of the Hyperarid Atacama Desert from Genome-Resolved Metagenomics." *Frontiers in Microbiology* 8. doi: ARTN 143510.3389/fmicb.2017.01435.

Fischer, S., Benz, J., Spath, B., Jellen-Ritter, A., Heyer, R., Dorr, M., … Marchfelder, A. (2011). Regulatory RNAs in Haloferax volcanii. *Biochem Soc Trans*, *39*(1), 159–162. https://doi.org/10.1042/BST0390159

Fischer, S., Benz, J., Spath, B., Maier, L. K., Straub, J., Granzow, M., … Marchfelder, A. (2010). The archaeal Lsm protein binds to small RNAs. *J Biol Chem*, *285*(45), 34429–34438. https://doi.org/10.1074/jbc.M110.118950

Forman, H. J., Zhang, H., & Rinna, A. (2009). Glutathione: overview of its protective roles, measurement, and biosynthesis. *Molecular Aspects of Medicine*, *30*(1–2), 1–

12. https://doi.org/10.1016/j.mam.2008.08.006

French, S. L., Santangelo, T. J., Beyer, A. L., & Reeve, J. N. (2007). Transcription and Translation are Coupled in Archaea. *Molecular Biology and Evolution*, *24*(4), 893–895. https://doi.org/10.1093/molbev/msm007

FRESNO, M., JIMÉNEZ, A., & VÁZQUEZ, D. (1977). Inhibition of Translation in Eukaryotic Systems by Harringtonine. *European Journal of Biochemistry*, *72*(2), 323–330. https://doi.org/10.1111/j.1432-1033.1977.tb11256.x

Gebetsberger, J., Wyss, L., Mleczko, A. M., Reuther, J., & Polacek, N. (2017). A tRNA-derived fragment competes with mRNA for ribosome binding and regulates translation during stress. *RNA Biol*, *14*(10), 1364–1373. https://doi.org/10.1080/15476286.2016.1257470

Gelsinger, D. R., and J. DiRuggiero. 2018a. "Transcriptional Landscape and Regulatory Roles of Small Noncoding RNAs in the Oxidative Stress Response of the Haloarchaeon Haloferax volcanii." *Journal of Bacteriology* 200 (9). doi: ARTN e00779-1710.1128/JB.00779-17.

Gelsinger, Diego, and Jocelyne DiRuggiero. 2018b. "The Non-Coding Regulatory RNA Revolution in Archaea." *Genes* 9 (3):141.

Genuth, N. R., & Barna, M. (2018). The Discovery of Ribosome Heterogeneity and Its Implications for Gene Regulation and Organismal Life. *Molecular Cell*, *71*(3), 364–374. https://doi.org/https://doi.org/10.1016/j.molcel.2018.07.018

Gerashchenko, M. V, & Gladyshev, V. N. (2014). Translation inhibitors cause abnormalities in ribosome profiling experiments. *Nucleic Acids Research*, *42*(17), e134–e134. https://doi.org/10.1093/nar/gku671

Gerashchenko, M. V, & Gladyshev, V. N. (2017). Ribonuclease selection for ribosome profiling. *Nucleic Acids Research*, *45*(2), e6–e6. https://doi.org/10.1093/nar/gkw822

Giles, N. M., Watts, A. B., Giles, G. I., Fry, F. H., Littlechild, J. A., & Jacob, C. (2003). Metal and Redox Modulation of Cysteine Protein Function. *Chemistry & Biology*, *10*(8), 677–693. https://doi.org/https://doi.org/10.1016/S1074-5521(03)00174-1

Gong, F., & Yanofsky, C. (2002). Instruction of Translating Ribosome by Nascent Peptide. *Science*, *297*(5588), 1864 LP – 1867. https://doi.org/10.1126/science.1073997

Gottesman, S., & Storz, G. (2011). Bacterial small RNA regulators: versatile roles and rapidly evolving variations. *Cold Spring Harb Perspect Biol*, *3*(12). https://doi.org/10.1101/cshperspect.a003798

Grabowski, B., & Kelman, Z. (2003). Archaeal DNA replication: eukaryal proteins in a bacterial context. *Annual Review of Microbiology*, *57*, 487–516. https://doi.org/10.1146/annurev.micro.57.030502.090709

Guo, H. (2018). Specialized ribosomes and the control of translation. *Biochemical Society Transactions*, *46*(4), 855–869. https://doi.org/10.1042/BST20160426

Halohandbook. Print.

He, Lin, and Gregory J. Hannon. "Micrornas: Small Rnas with a Big Role in Gene Regulation." *Nat Rev Genet* 5.7 (2004): 522-31. Print.

Hendrickson, Erik L., et al. "Global Responses of Methanococcus Maripaludis to Specific Nutrient Limitations and Growth Rate." *Journal of Bacteriology* 190.6 (2008): 2198-205. Print.

Heyer, R., Dorr, M., Jellen-Ritter, A., Spath, B., Babski, J., Jaschinski, K., …

Marchfelder, A. (2012). High throughput sequencing reveals a plethora of small RNAs including tRNA derived fragments in Haloferax volcanii. *RNA Biol*, *9*(7), 1011–1018. https://doi.org/10.4161/rna.20826

Hinnebusch, A. G., & Lorsch, J. R. (2012). The Mechanism of Eukaryotic Translation Initiation: New Insights and Challenges. *Cold Spring Harbor Perspectives in Biology* , *4*(10). https://doi.org/10.1101/cshperspect.a011544

Hoffmann, Lena, et al. "Expanding the Archaellum Regulatory Network – the Eukaryotic Protein Kinases Arnc and Arnd Influence Motility of Sulfolobus Acidocaldarius." *MicrobiologyOpen* 6.1 (2017): e00414. Print.

Hwang, J.-Y., & Buskirk, A. R. (2017). A ribosome profiling study of mRNA cleavage by the endonuclease RelE. *Nucleic Acids Research*, *45*(1), 327–336. https://doi.org/10.1093/nar/gkw944

Imachi, H., Nobu, M. K., Nakahara, N., Morono, Y., Ogawara, M., Takaki, Y., … Takai, K. (2019). Isolation of an archaeon at the prokaryote-eukaryote interface. *BioRxiv*. https://doi.org/10.1101/726976

Imlay, J. A. (2003). Pathways of oxidative damage. *Annu Rev Microbiol*, *57*, 395–418. https://doi.org/10.1146/annurev.micro.57.030502.090938

Imlay, J. A. (2006). Iron-sulphur clusters and the problem with oxygen. *Mol Microbiol*, *59*(4), 1073–1082. https://doi.org/MMI5028 [pii]10.1111/j.1365-2958.2006.05028.x

Imlay, J. A. (2008). Cellular defenses against superoxide and hydrogen peroxide. *Annu Rev Biochem*, *77*, 755–776. https://doi.org/10.1146/annurev.biochem.77.061606.161055

Ingolia, N. T., Lareau, L. F., & Weissman, J. S. (2011). Ribosome Profiling of Mouse

Embryonic Stem Cells Reveals the Complexity and Dynamics of Mammalian Proteomes. *Cell*, *147*(4), 789–802. https://doi.org/https://doi.org/10.1016/j.cell.2011.10.002

Jäger, D., Förstner, K. U., Sharma, C. M., Santangelo, T. J., & Reeve, J. N. (2014). Primary transcriptome map of the hyperthermophilic archaeon Thermococcus kodakarensis. *BMC Genomics*, *15*(1), 684. https://doi.org/10.1186/1471-2164-15-684

Jager, D., Pernitzsch, S. R., Richter, A. S., Backofen, R., Sharma, C. M., & Schmitz, R. A. (2012). An archaeal sRNA targeting cis- and trans-encoded mRNAs via two distinct domains. *Nucleic Acids Res*, *40*(21), 10964–10979. https://doi.org/10.1093/nar/gks847

Jager, D., Sharma, C. M., Thomsen, J., Ehlers, C., Vogel, J., & Schmitz, R. A. (2009). Deep sequencing analysis of the Methanosarcina mazei Go1 transcriptome in response to nitrogen availability. *Proc Natl Acad Sci U S A*, *106*(51), 21878–21882. https://doi.org/10.1073/pnas.0909051106

Jarrell, K. F., Walters, A. D., Bochiwal, C., Borgia, J. M., Dickinson, T., & Chong, J. P. J. (2011). Major players on the microbial stage: why archaea are important. *Microbiology*, *157*(4), 919–936. https://doi.org/https://doi.org/10.1099/mic.0.047837-0

Jaschinski, K., Babski, J., Lehr, M., Burmester, A., Benz, J., Heyer, R., … Soppa, J. (2014). Generation and Phenotyping of a Collection of sRNA Gene Deletion Mutants of the Haloarchaeon Haloferax volcanii. *PLOS One*, *9*(3), e90763. https://doi.org/10.1371/journal.pone.0090763

Jeong, Y., Kim, J.-N., Kim, M. W., Bucca, G., Cho, S., Yoon, Y. J., … Cho, B.-K. (2016). The dynamic transcriptional and translational landscape of the model antibiotic producer Streptomyces coelicolor A3(2). *Nature Communications*, *7*(1), 11605. https://doi.org/10.1038/ncomms11605

Jiang, X., Keto-Timonen, R., Skurnik, M., & Korkeala, H. (2019). Role of DEAD-box RNA helicase genes in the growth of Yersinia pseudotuberculosis IP32953 under cold, pH, osmotic, ethanol and oxidative stresses. *PLOS ONE*, *14*(7), e0219422.

Kalvari, Ioanna, Joanna Argasinska, Natalia Quinones-Olvera, Eric P. Nawrocki, Elena Rivas, Sean R. Eddy, Alex Bateman, Robert D. Finn, and Anton I. Petrov. 2017. "Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families." *Nucleic Acids Research* 46 (D1):D335-D342. doi: 10.1093/nar/gkx1038.

Karlsen, J., Asplund-Samuelsson, J., Thomas, Q., Jahn, M., & Hudson, E. P. (2018). Ribosome Profiling of Synechocystis Reveals Altered Ribosome Allocation at Carbon Starvation. *MSystems*, *3*(5), e00126-18. https://doi.org/10.1128/mSystems.00126-18

Kaur, A., et al. "Coordination of Frontline Defense Mechanisms under Severe Oxidative Stress." *Mol Syst Biol* 393 (2010): doi:10.1038/msb.2010.50. Print.

Kim, Daehwan, Ben Langmead, and Steven L. Salzberg. "Hisat: A Fast Spliced Aligner with Low Memory Requirements." *Nat Meth* 12.4 (2015): 357-60. Print.

Kish, A., G. Kirkali, C. Robinson, R. Rosenblatt, P. Jaruga, M. Dizdaroglu, and J. DiRuggiero. 2009. "Salt shield: intracellular salts provide cellular protection against ionizing radiation in the halophilic archaeon, Halobacterium salinarum NRC-1." *Environmental microbiology* 11 (5):1066.

Kliemt, Jana, Katharina Jaschinski, and Jörg Soppa. 2019. "A Haloarchaeal Small Regulatory RNA (sRNA) Is Essential for Rapid Adaptation to Phosphate Starvation Conditions." *Frontiers in microbiology* 10:1219-1219. doi: 10.3389/fmicb.2019.01219.

Kramer, P., Gäbel, K., Pfeiffer, F., & Soppa, J. (2014). Haloferax volcanii, a prokaryotic species that does not use the Shine Dalgarno mechanism for translation initiation at 5'-UTRs. *PloS One*, *9*(4), e94979–e94979. https://doi.org/10.1371/journal.pone.0094979

Krulwich, Terry A., David B. Hicks, and Masahiro Ito. 2009. "Cation/proton antiporter complements of bacteria: why so large and diverse?" *Molecular microbiology* 74 (2):257-260. doi: 10.1111/j.1365-2958.2009.06842.x.

Laass, S., Monzon, V. A., Kliemt, J., Hammelmann, M., Pfeiffer, F., Förstner, K. U., & Soppa, J. (2019). Characterization of the transcriptome of Haloferax volcanii, grown under four different conditions, with mixed RNA-Seq. *PLOS ONE*, *14*(4), e0215986. Retrieved from https://doi.org/10.1371/journal.pone.0215986

Lalaouna, D., Prevost, K., Eyraud, A., & Masse, E. (2017). Identification of unknown RNA partners using MAPS. *Methods (San Diego, Calif.)*, *117*, 28–34. https://doi.org/10.1016/j.ymeth.2016.11.011

Lambrecht, Jennifer A., Jeffrey M. Flynn, and Diana M. Downs. "Conserved Yjgf Protein Family Deaminates Reactive Enamine/Imine Intermediates of Pyridoxal 5′-Phosphate (Plp)-Dependent Enzyme Reactions." *The Journal of Biological Chemistry* 287.5 (2012): 3454-61. Print.

Lecompte, O., Ripp, R., Thierry, J., Moras, D., & Poch, O. (2002). Comparative analysis

of ribosomal proteins in complete genomes: an example of reductive evolution at the domain scale. *Nucleic Acids Research*, *30*(24), 5382–5390. https://doi.org/10.1093/nar/gkf693

Lee, S., Liu, B., Lee, S., Huang, S.-X., Shen, B., & Qian, S.-B. (2012). Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. *Proceedings of the National Academy of Sciences*, *109*(37), E2424 LP-E2432. https://doi.org/10.1073/pnas.1207846109

Levy, S., Portnoy, V., Admon, J., & Schuster, G. (2011). Distinct activities of several RNase J proteins in methanogenic archaea. *RNA Biology*, *8*(6), 1073–1083. https://doi.org/10.4161/rna.8.6.16604

Li, Bo, and Colin N. Dewey. "Rsem: Accurate Transcript Quantification from Rna-Seq Data with or without a Reference Genome." *BMC Bioinformatics* 12 (2011): 323-23. Print.

Li, Bo, et al. "Rna-Seq Gene Expression Estimation with Read Mapping Uncertainty." *Bioinformatics* 26.4 (2010): 493-500. Print.

Li, J., Qi, L., Guo, Y., Yue, L., Li, Y., Ge, W., … Dong, X. (2015). Global mapping transcriptional start sites revealed both transcriptional and post-transcriptional regulation of cold adaptation in the methanogenic archaeon Methanolobus psychrophilus. *Scientific Reports*, *5*, 9209. https://doi.org/10.1038/srep09209

Li, Y., Liu, X., Huang, L., Guo, H., & Wang, X.-J. (2010). Potential coexistence of both bacterial and eukaryotic small RNA biogenesis and functional related protein homologs in Archaea. *Journal of Genetics and Genomics = Yi Chuan Xue Bao*, *37*(8), 493–503. https://doi.org/10.1016/S1673-8527(09)60069-2

Liao, Yang, Gordon K. Smyth, and Wei Shi. "Featurecounts: An Efficient General Purpose Program for Assigning Sequence Reads to Genomic Features." *Bioinformatics* 30.7 (2014): 923-30. Print.

Lindas, Ann-Christin, and Rolf Bernander. "The Cell Cycle of Archaea." *Nat Rev Micro* 11.9 (2013): 627-38. Print.

Lloréns-Rico, Verónica, Jaime Cano, Tjerko Kamminga, Rosario Gil, Amparo Latorre, Wei-Hua Chen, Peer Bork, John I. Glass, Luis Serrano, and Maria Lluch-Senar. 2016. "Bacterial antisense RNAs are mainly the product of transcriptional noise." *Science advances* 2 (3):e1501363-e1501363. doi: 10.1126/sciadv.1501363.

Love, M. I., W. Huber, and S. Anders. 2014. "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2." *Genome Biol* 15 (12):550. doi: 10.1186/s13059-014-0550-8.

Mack, George S. "Microrna Gets Down to Business." *Nat Biotech* 25.6 (2007): 631-38. Print.

Mann, Martin, Patrick R. Wright, and Rolf Backofen. 2017. "IntaRNA 2.0: enhanced and customizable prediction of RNA–RNA interactions." *Nucleic Acids Research* 45 (W1):W435-W439. doi: 10.1093/nar/gkx279.

Marchfelder, A., Fischer, S., Brendel, J., Stoll, B., Maier, L. K., Jager, D., … Randau, L. (2012). Small RNAs for defence and regulation in archaea. *Extremophiles*, *16*(5), 685–696. https://doi.org/10.1007/s00792-012-0469-5

Märtens, B., Manoharadas, S., Hasenöhrl, D., Manica, A., & Bläsi, U. (2013). Antisense regulation by transposon-derived RNAs in the hyperthermophilic archaeon <em>Sulfolobus solfataricus</em&gt. *EMBO Reports*, *14*(6), 527.

Retrieved from http://embor.embopress.org/content/14/6/527.abstract

McCready, S., et al. "Uv Irradiation Induces Homologous Recombination Genes in the Model Archaeon, Halobacterium Sp. Nrc-1." *Saline Systems* 1 (2005): 3.

McMillan, L. J., Hwang, S., Farah, R. E., Koh, J., Chen, S., & Maupin-Furlow, J. A. (2018). Multiplex quantitative SILAC for analysis of archaeal proteomes: a case study of oxidative stress responses. *Environmental Microbiology*, *20*(1), 385–401. https://doi.org/10.1111/1462-2920.14014

Meslier, V., M.C. Casero, M. Daily, J. Wierchos, C. Ascaso, O. Artieda, P.R. McCullough, and J. DiRuggiero. 2018. "Fundamental drivers for endolithic microbial community assemblies in the hyperarid Atacama Desert." *Env. Microbiol.* 20:1765-1781.

Meydan, S., Marks, J., Klepacki, D., Sharma, V., Baranov, P. V, Firth, A. E., Mankin, A. S. (2019). Retapamulin-Assisted Ribosome Profiling Reveals the Alternative Bacterial Proteome. *Molecular Cell*, *74*(3), 481-493.e6. https://doi.org/https://doi.org/10.1016/j.molcel.2019.02.017

Mohammad, F., Green, R., & Buskirk, A. R. (2019). A systematically-revised ribosome profiling method for bacteria reveals pauses at single-codon resolution. *ELife*, *8*, e42591. https://doi.org/10.7554/eLife.42591

Mohammad, F., Woolstenhulme, C. J., Green, R., & Buskirk, A. R. (2016). Clarifying the Translational Pausing Landscape in Bacteria by Ribosome Profiling. *Cell Reports*, *14*(4), 686–694. https://doi.org/https://doi.org/10.1016/j.celrep.2015.12.073

Molina-Henares, Antonio J., Tino Krell, Maria Eugenia Guazzaroni, Ana Segura, and Juan L. Ramos. 2006. "Members of the IclR family of bacterial transcriptional regulators function as activators and/or repressors." *FEMS Microbiology Reviews*

30 (2):157-186. doi: 10.1111/j.1574-6976.2005.00008.x.

Morris, K. V, & Mattick, J. S. (2014). The rise of regulatory RNA. *Nature Reviews. Genetics*, *15*(6), 423–437. https://doi.org/10.1038/nrg3722

Mountfort, D. O. (1978). Evidence for ATP synthesis driven by a proton gradient in Methanosarcinabarkeri. *Biochemical and Biophysical Research Communications*, *85*(4), 1346–1351. https://doi.org/https://doi.org/10.1016/0006-291X(78)91151-8

Mukhopadhyay, Biswarup, Eric F. Johnson, and Ralph S. Wolfe. "A Novel Ph2 Control on the Expression of Flagella in the Hyperthermophilic Strictly Hydrogenotrophic Methanarchaeaon Methanococcus Jannaschii." *Proceedings of the National Academy of Sciences* 97.21 (2000): 11522-27. Print.

Müller, Alexandra, et al. "Activation of Rida Chaperone Function by N-Chlorination." *Nature Communications* 5 (2014): 5804. Print.

Orell, A., Tripp, V., Aliaga-Tobar, V., Albers, S.-V., Maracaja-Coutinho, V., & Randau, L. (2018). A regulatory RNA is involved in RNA duplex formation and biofilm regulation in Sulfolobus acidocaldarius. *Nucleic Acids Research*, *46*(9), 4794–4806. https://doi.org/10.1093/nar/gky144

Oren, A. (2014). Halophilic archaea on Earth and in space: growth and survival under extreme conditions. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *372*(2030), 20140194. https://doi.org/10.1098/rsta.2014.0194

Nakahigashi, K., Takai, Y., Shiwa, Y., Wada, M., Honma, M., Yoshikawa, H., … Mori, H. (2014). Effect of codon adaptation on codon-level and gene-level translation efficiency in vivo. *BMC Genomics*, *15*(1), 1115. https://doi.org/10.1186/1471-2164-

15-1115

Nakatogawa, H., & Ito, K. (2002). The Ribosomal Exit Tunnel Functions as a Discriminating Gate. *Cell*, *108*(5), 629–636. https://doi.org/https://doi.org/10.1016/S0092-8674(02)00649-9

Nitzan, Mor, Rotem Rehani, and Hanah Margalit. 2017. "Integration of Bacterial Small RNAs in Regulatory Networks." *Annual Review of Biophysics* 46 (1):131-148. doi: 10.1146/annurev-biophys-070816-034058.

Nyström, Thomas. "Role of Oxidative Carbonylation in Protein Quality Control and Senescence." *The EMBO Journal* 24.7 (2005): 1311-17. Print.

Oren, A. (2013). Life at high salt concentrations, intracellular KCl concentrations, and acidic proteomes. *Frontiers in Microbiology*, *4*, 315. https://doi.org/10.3389/fmicb.2013.00315

Peeters, E., Driessen, R. P. C., Werner, F., & Dame, R. T. (2015). The interplay between nucleoid organization and transcription in archaeal genomes. *Nature Reviews Microbiology*, *13*, 333.

Perez-Rueda, Ernesto, Rafael Hernandez-Guerrero, Mario Alberto Martinez-Nuñez, Dagoberto Armenta-Medina, Israel Sanchez, and J. Antonio Ibarra. 2018. "Abundance, diversity and domain architecture variability in prokaryotic DNA-binding transcription factors." *PloS one* 13 (4):e0195332-e0195332. doi: 10.1371/journal.pone.0195332.

Pertea, Mihaela, et al. "Transcript-Level Expression Analysis of Rna-Seq Experiments with Hisat, Stringtie and Ballgown." *Nat. Protocols* 11.9 (2016): 1650-67. Print.

Pertea, Mihaela, et al. "Stringtie Enables Improved Reconstruction of a Transcriptome from Rna-Seq Reads." *Nat Biotech* 33.3 (2015): 290-95. Print.

Pohlschroder, M., & Schulze, S. (2019). Haloferax volcanii. *Trends in Microbiology*, *27*(1), 86–87. https://doi.org/https://doi.org/10.1016/j.tim.2018.10.004

Pointing, Stephen B., and Jayne Belnap. 2012. "Microbial colonization and controls in dryland systems." *Nature Reviews Microbiology* 10:551. doi: 10.1038/nrmicro2831.

Prasse, D, Ehlers, C., Backofen, R., & Schmitz, R. A. (2013). Regulatory RNAs in archaea: first target identification in Methanoarchaea. *Biochem Soc Trans*, *41*(1), 344–349. https://doi.org/10.1042/BST20120280

Prasse, Daniela, Förstner, K. U., Jäger, D., Backofen, R., & Schmitz, R. A. (2017). sRNA154 a newly identified regulator of nitrogen fixation in Methanosarcina mazei strain Gö1. *RNA Biol*, *14*(11), 1544–1558. https://doi.org/10.1080/15476286.2017.1306170

Preiss, T. (2016). All Ribosomes Are Created Equal. Really? *Trends in Biochemical Sciences*, *41*(2), 121–123. https://doi.org/https://doi.org/10.1016/j.tibs.2015.11.009

Raman, Arjun V. and López García de Lomana, Adrián and Kusebauch, Ulrike and Pan, Min and Turkarslan, Serdar and Moritz, Robert L. and Baliga, N. S. (2019). Context-Specific Regulation of Coupled Transcription-Translation Modules Predicts Pervasive Ribosome Specialization. *Nature Structural Biology*.

Randau, L. (2012). RNA processing in the minimal organism Nanoarchaeum equitans. *Genome Biol*, *13*(7), R63. https://doi.org/10.1186/gb-2012-13-7-r63

Robert, F., Carrier, M., Rawe, S., Chen, S., Lowe, S., & Pelletier, J. (2009). Altering

Chemosensitivity by Modulating Translation Elongation. *PLOS ONE*, *4*(5), e5428.

Robinson, C. K., J. Wierzchos, C. Black, A. Crits-Christoph, B. Ma, J. Ravel, C. Ascaso, O. Artieda, S. Valea, M. Roldan, B. Gomez-Silva, and J. DiRuggiero. 2015. "Microbial diversity and the presence of algae in halite endolithic communities are correlated to atmospheric moisture in the hyper-arid zone of the Atacama Desert." *Environ Microbiol* 17:299-315. doi: 10.1111/1462-2920.12364.

Robinson, C. K., Webb, K., Kaur, A., Jaruga, P., Dizdaroglu, M., Baliga, N. S., … Diruggiero, J. (2011). A major role for nonenzymatic antioxidant processes in the radioresistance of Halobacterium salinarum. *J Bacteriol*, *193*(7), 1653–1662. https://doi.org/10.1128/JB.01310-10

Robinson, James T., et al. "Integrative Genomics Viewer." *Nat Biotech* 29.1 (2011): 24-26. Print.

Scharf, Birgit, and Martin Engelhard. "Halocyanin, an Archaebacterial Blue Copper Protein (Type I) from Natronobacterium Pharaonis." *Biochemistry* 32.47 (1993): 12894-900. Print.

Santangelo, T. J., Cubonova, L., Matsumi, R., Atomi, H., Imanaka, T., & Reeve, J. N. (2008). Polarity in Archaeal Operon Transcription in Thermococcus kodakaraensis. *J. BActeriol.*, *190*(6), 2244–2248. https://doi.org/10.1128/jb.01811-07

Schmitz-Streit, R., Jäger, D., Jellen-Ritter, A., Babski, J., Soppa, J., & Marchfelder, A. (2011). Archaea employ small RNAs as regulators. In W. R. Hess & A. Marchfelder (Eds.), *Regulatory RNAs in Prokaryotes* (pp. 131–145). Wien: Springer Verlag.

Sharma, C. M., & Vogel, J. (2014). Differential RNA-seq: the approach behind and the biological insight gained. *Current Opinion in Microbiology*, *19*, 97–105.

https://doi.org/https://doi.org/10.1016/j.mib.2014.06.010

Sharma, K., et al. "The Rosr Transcription Factor Is Required for Gene Expression Dynamics in Response to Extreme Oxidative Stress in a Hypersaline-Adapted Archaeon." *BMC Genomics* 13 (2012): 351. Print.

Shell, S. S., Wang, J., Lapierre, P., Mir, M., Chase, M. R., Pyle, M. M., … Gray, T. A. (2015). Leaderless Transcripts and Small Proteins Are Common Features of the Mycobacterial Translational Landscape. *PLOS Genetics*, *11*(11), e1005641. Retrieved from https://doi.org/10.1371/journal.pgen.1005641

Shi, Y. M., G. W. Tyson, and E. F. DeLong. 2009. "Metatranscriptomics reveals unique microbial small RNAs in the ocean's water column." *Nature* 459 (7244):266-U154. doi: 10.1038/nature08055.

Skowyra, A., and S. A. MacNeill. "Identification of Essential and Non-Essential Single-Stranded DNA-Binding Proteins in a Model Archaeal Organism." *Nucleic Acids Res* 40.3 (2012): 1077-90. Print.

Smirnov, A., Förstner, K. U., Holmqvist, E., Otto, A., Günster, R., Becher, D., … Vogel, J. (2016). Grad-seq guides the discovery of ProQ as a major small RNA-binding protein. *Proceedings of the National Academy of Sciences*, *113*(41), 11591–11596. https://doi.org/10.1073/pnas.1609981113

Soppa, J., Straub, J., Brenneis, M., Jellen-Ritter, A., Heyer, R., Fischer, S., … Marchfelder, A. (2009). Small RNAs of the halophilic archaeon Haloferax volcanii. *Biochem Soc Trans*, *37*(Pt 1), 133–136. https://doi.org/10.1042/BST0370133

Spang, A., Saw, J. H., Jørgensen, S. L., Zaremba-Niedzwiedzka, K., Martijn, J., Lind, A. E., … Ettema, T. J. G. (2015). Complex archaea that bridge the gap between

prokaryotes and eukaryotes. *Nature*, *521*, 173.

Stadtman, E. R. (1993). Oxidation of free amino acids and amino acid residues in proteins by radiolysis and by metal-catalyzed reactions. *Annu Rev Biochem*, *62*, 797–821. https://doi.org/10.1146/annurev.bi.62.070193.004053

Storz, G., Vogel, J., & Wassarman, K. M. (2011). Regulation by small RNAs in bacteria: expanding frontiers. *Mol Cell*, *43*(6), 880–891. https://doi.org/10.1016/j.molcel.2011.08.022

Straub, J., Brenneis, M., Jellen-Ritter, A., Heyer, R., Soppa, J., & Marchfelder, A. (2009). Small RNAs in haloarchaea: identification, differential expression and biological function. *RNA Biol*, *6*(3), 281–292. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/19333006

Stroud, A., S. Liddell, and T. Allers. "Genetic and Biochemical Identification of a Novel Single-Stranded DNA-Binding Complex in Haloferax Volcanii." *Front Microbiol* 3 (2012): 224. Print.

Stutz, Jochen, et al. "Atmospheric Reactive Chlorine and Bromine at the Great Salt Lake, Utah." *Geophysical Research Letters* 29.10 (2002): 18-1-18-4. Print.

Suzuki, Yuichiro J., Marina Carini, and D. Allan Butterfield. "Protein Carbonylation." *Antioxidants & Redox Signaling* 12.3 (2010): 323-25. Print.

Takara, T. J., and S. P. Bell. 2009. "Putting two heads together to unwind DNA." *Cell* 139 (4):652-654. doi: S0092-8674(09)01363-4 [pii]10.1016/j.cell.2009.10.037.

Thomason, M. K., & Storz, G. (2010). Bacterial antisense RNAs: How many are there and what are they doing? *Annual Review of Genetics*, *44*, 167–188. https://doi.org/10.1146/annurev-genet-102209-163523

Toffano-Nioche, C., Ott, A., Crozat, E., Nguyen, A. N., Zytnicki, M., Leclerc, F., … Gautheret, D. (2013). RNA at 92°C: The non-coding transcriptome of the hyperthermophilic archaeon Pyrococcus abyssi. *RNA Biol*, *10*(7), 1211–1220. https://doi.org/10.4161/rna.25567

Tosa, T., & Pizer, L. I. (1971). Biochemical Bases for the Antimetabolite Action of l-Serine Hydroxamate. *Journal of Bacteriology*, *106*(3), 972 LP – 982. Retrieved from http://jb.asm.org/content/106/3/972.abstract

Toyofuku, Masanori, Nobuhiko Nomura, and Leo Eberl. 2019. "Types and origins of bacterial membrane vesicles." *Nature Reviews Microbiology* 17 (1):13-24. doi: 10.1038/s41579-018-0112-2.

Tripepi, Manuela, Saheed Imam, and Mechthild Pohlschröder. "Haloferax Volcanii Flagella Are Required for Motility but Are Not Involved in Pibd-Dependent Surface Adhesion." *Journal of Bacteriology* 192.12 (2010): 3093-102. Print.

Tripepi, Manuela, et al. "N-Glycosylation of Haloferax Volcanii Flagellins Requires Known Agl Proteins and Is Essential for Biosynthesis of Stable Flagella." *Journal of Bacteriology* 194.18 (2012): 4876-87. Print.

Tsatsaronis, James A., Sandra Franch-Arroyo, Ulrike Resch, and Emmanuelle Charpentier. 2018. "Extracellular Vesicle RNA: A Universal Mediator of Microbial Communication?" *Trends in Microbiology* 26 (5):401-410. doi: 10.1016/j.tim.2018.02.009.

Uritskiy, G.V., and J. DiRuggiero. 2019. "Applying Genome-Resolved Metagenomics to Deconvolute the Halophilic Microbiome." *Gene* 10:220.

Uritskiy, Gherman, Samantha Getsin, Adam Munn, Benito Gomez-Silva, Alfonso Davila, Brian Glass, James Taylor, and Jocelyne DiRuggiero. 2019. "Halophilic microbial community compositional shift after a rare rainfall in the Atacama Desert." *The ISME Journal*. doi: 10.1038/s41396-019-0468-y.

Uritskiy, Gherman V., Jocelyne DiRuggiero, and James Taylor. 2018. "MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis." *Microbiome* 6 (1):158. doi: 10.1186/s40168-018-0541-1.

Villanueva, L., Damsté, J. S. S., & Schouten, S. (2014). A re-evaluation of the archaeal membrane lipid biosynthetic pathway. *Nature Reviews Microbiology*, *12*, 438.

Wagner, Günter P., Koryu Kin, and Vincent J. Lynch. "Measurement of Mrna Abundance Using Rna-Seq Data: Rpkm Measure Is Inconsistent among Samples." *Theory in Biosciences* 131.4 (2012): 281-85. Print.

Wagner, E. G. H., & Romby, P. (2015). Chapter Three - Small RNAs in Bacteria and Archaea: Who They Are, What They Do, and How They Do It. In T. Friedmann, J. C. Dunlap, & S. F. Goodwin (Eds.), *Advances in Genetics* (Vol. 90, pp. 133–208). https://doi.org/https://doi.org/10.1016/bs.adgen.2015.05.001

Wassarman, Karen M. 2018. "6S RNA, a Global Regulator of Transcription." *Microbiology spectrum* 6 (3):10.1128/microbiolspec.RWR-0019-2018. doi: 10.1128/microbiolspec.RWR-0019-2018.

Waters, L. S., and G. Storz. "Regulatory Rnas in Bacteria." *Cell* 136.4 (2009): 615-28. Print.

Weaver, J., Mohammad, F., Buskirk, A. R., & Storz, G. (2019). Identifying Small Proteins by Ribosome Profiling with Stalled Initiation Complexes. *MBio*, *10*(2).

https://doi.org/10.1128/mBio.02819-18

Webb, K.M., et al. "Effects of Intracellular Mn on the Radiation Resistance of the Halophilic Archaeon Halobacterium Salinarum." *Extremophiles* 17 (2013): 485-97. Print.

Weidmann, C. A., Mustoe, A. M., & Weeks, K. M. (2016). Direct Duplex Detection: An Emerging Tool in the RNA Structure Analysis Toolbox. *Trends in Biochemical Sciences*, *41*(9), 734–736. https://doi.org/10.1016/j.tibs.2016.07.001

Weinberg, Zasha, Christina E. Lünse, Keith A. Corbino, Tyler D. Ames, James W. Nelson, Adam Roth, Kevin R. Perkins, Madeline E. Sherlock, and Ronald R. Breaker. 2017. "Detection of 224 candidate structured RNAs by comparative analysis of specific subsets of intergenic regions." *Nucleic Acids Research* 45 (18):10811-10823. doi: 10.1093/nar/gkx699.

Wheeler, Bayly S. "Small Rnas, Big Impact: Small Rna Pathways in Transposon Control and Their Effect on the Host Stress Response." *Chromosome Research* 21.6 (2013): 587-600. Print.

Whitehead, K., Kish, A., Pan, M., Kaur, A., Reiss, D. J., King, N., Baliga, N. S. (2006). An integrated systems approach for understanding cellular responses to gamma radiation. *Mol Syst Biol*, *47*, doi:10.1038/msb4100091.

Will, Sebastian, Tejal Joshi, Ivo L. Hofacker, Peter F. Stadler, and Rolf Backofen. 2012. "LocARNA-P: Accurate boundary prediction and improved detection of structural RNAs." *RNA* 18 (5):900-914.

Williams, T. A., Foster, P. G., Cox, C. J., & Embley, T. M. (2013). An archaeal origin of eukaryotes supports only two primary domains of life. *Nature*, *504*, 231.

Willkomm, Sarah, et al. "Structural and Mechanistic Insights into an Archaeal DNA-Guided Argonaute Protein." *Nature Microbiology* 2 (2017): 17035. Print.

Willkomm, Sarah, et al. "A Prokaryotic Twist on Argonaute Function." *Life* 5.1 (2015): 538. Print.

Woese, C. (1998). The universal ancestor. *Proceedings of the National Academy of Sciences*, *95*(12), 6854–6859. https://doi.org/10.1073/PNAS.95.12.6854

Woolstenhulme, C. J., Guydosh, N. R., Green, R., & Buskirk, A. R. (2015). High-Precision Analysis of Translational Pausing by Ribosome Profiling in Bacteria Lacking EFP. *Cell Reports*, *11*(1), 13–21. https://doi.org/https://doi.org/10.1016/j.celrep.2015.03.014

Wu, C. C.-C., Zinshteyn, B., Wehner, K. A., & Green, R. (2019). High-Resolution Ribosome Profiling Defines Discrete Ribosome Elongation States and Translational Regulation during Cellular Stress. *Molecular Cell*, *73*(5), 959-970.e5. https://doi.org/https://doi.org/10.1016/j.molcel.2018.12.009

Wurtmann, E. J., Ratushny, A. V, Pan, M., Beer, K. D., Aitchison, J. D., & Baliga, N. S. (2014). An evolutionarily conserved RNase-based mechanism for repression of transcriptional positive autoregulation. *Molecular Microbiology*, *92*(2), 369–382. https://doi.org/10.1111/mmi.12564

Wurtzel, O., et al. "A Single-Base Resolution Map of an Archaeal Transcriptome." *Genome Res* 20.1 (2010): 133-41. Print.

Wyss, L., Waser, M., Gebetsberger, J., Zywicki, M., & Polacek, N. (2018). mRNA-specific translation regulation by a ribosome-associated ncRNA in Haloferax volcanii. *Scientific Reports*, *8*(1), 12502. https://doi.org/10.1038/s41598-018-30332-

w

Xia, Qiangwei, et al. "Quantitative Proteomics of Nutrient Limitation in the Hydrogenotrophic Methanogen Methanococcus Maripaludis." *BMC Microbiology* 9 (2009): 149-49. Print.

Xiao, W., Wang, R.-S., Handy, D. E., & Loscalzo, J. (2018). NAD(H) and NADP(H) Redox Couples and Cellular Energy Metabolism. *Antioxidants & Redox Signaling*, *28*(3), 251–272. https://doi.org/10.1089/ars.2017.7216

Xue, S., & Barna, M. (2012). Specialized ribosomes: a new frontier in gene regulation and organismal biology. *Nature Reviews Molecular Cell Biology*, *13*(6), 355–369. https://doi.org/10.1038/nrm3359

Yanofsky, C. (1981). Attenuation in the control of expression of bacterial operons. *Nature*, *289*(5800), 751–758. https://doi.org/10.1038/289751a0

Yue, L., Li, J., Zhang, B., Qi, L., Zhao, F., Li, L., & Dong, X. (2019). aCPSF1 controlled archaeal transcription termination: a prototypical eukaryotic model. *BioRxiv*, 843821. https://doi.org/10.1101/843821

Zander, Adrian, et al. "Guide-Independent DNA Cleavage by Archaeal Argonaute from Methanocaldococcus Jannaschii." *Nature Microbiology* 2 (2017): 17034. Print.

Zhang, Y., Burkhardt, D. H., Rouskin, S., Li, G.-W., Weissman, J. S., & Gross, C. A. (2018). A Stress Response that Monitors and Regulates mRNA Structure Is Central to Cold Shock Adaptation. *Molecular Cell*, *70*(2), 274-286.e7. https://doi.org/10.1016/j.molcel.2018.02.035

Zhang, Zhao, et al. "Strand-Specific Libraries for High Throughput Rna Sequencing (Rna-Seq) Prepared without Poly(a) Selection." *Silence* 3.1 (2012): 9. Print.

VITA

DIEGO RIVERA GELSINGER

Department of Biology • Johns Hopkins University • Baltimore, MD 21218
dgelsin1@jhu.edu • (323) 314-4077
https://github.com/dgelsin
https://dgelsin.github.io

PROFESSIONAL STATUS

Johns Hopkins University - Baltimore, MD
2014–present
PhD Candidate, Department of Biology, Program in Cellular, Molecular, Developmental, and Biophysics (CMDB)
Advisor: Dr. Jocelyne DiRuggiero
Research focus: Mechanistic insight into small non-coding RNA (sRNA)-mediated regulation of stress response in halophilic archaea from desiccated environments and metatranscriptomic discovery of sRNAs in halophilic microbial communities from the Atacama Desert. High throughput translation dynamics in Archaea.

San Francisco State University - San Francisco, CA
2012–2014
Undergraduate Research Fellow, Department of Biology
Advisor: Dr. José de la Torre.
Research focus*:* Physiology and ecological interactions of thermophilic archaea and bacteria in geothermal hot springs across the globe.

EDUCATION

John Hopkins University - Baltimore, MD *Fall*
2014–May 2020
Ph.D. Microbiology

San Francisco State University - San Francisco, CA *Completed*
Spring 2014
B.Sc. Microbiology, minor in French

PUBLICATIONS

- **Gelsinger, D.R.***, Uritskiy, G.*, Reddy, R., Munn, A., Farney, S.K., and DiRuggiero, J. **2020**. "*Regulatory non-coding RNAs are diverse and abundant in an extremophilic microbial community".* mSystems.
- **Gelsinger, D.R.** and J. DiRuggiero. **2018**. "*The non-coding regulatory RNA revolution in Archaea*". Genes; 9(3), 141 https://doi.org/10.3390/genes9030141
- **Gelsinger, D.R.** and J. DiRuggiero. **2018**. "*Transcriptional landscape and regulatory roles of small non-coding RNAs in the oxidative stress response of the haloarchaeon Haloferax volcanii*". J Bacteriol; 200:e00779-17 doi: 10.1128/JB.00779-17

- Crits-Christoph, A., **Gelsinger, D.R.**, Ma, B., Wierzchos, J., Ravel, J., Davila, A.F., Casero, M.C., and DiRuggiero, J. **2016**. "*Functional interactions of archaea, bacteria and viruses in a hypersaline endolithic community*". Env Microbiol; 18:2064 PMID: 26914534
- Davila, A.F., Hawes, I., Araya, J.G., **Gelsinger**, **D.R.**, DiRuggiero, J., Ascaso, C., Osano, A., and Wierzchos, J. **2015**. "*In situ metabolism in halite endolithic microbial communities of the hyperarid Atacama Desert*". Front. Microbiol; 6:1035 PMID: 26500612
- Trujillo, G., Aguinaldo, P., Anderson, C., Bustamante, J., **Gelsinger, D.R.**, Pastor, M.J, Wright, J., Marquez-Magaña, L., and Riggs, B. **2015** "*Near-peer STEM Mentoring Offers unexpected benefits for mentors from traditionally underrepresented backgrounds*". *PURM.*

Under review

- **Gelsinger, D.R.**, Dallon, E., Reddy, R., Mohammed, F., Green, R., Buskirk, A., and DiRuggiero, J. **2020**. "*Ribosome profiling in Archaea reveals leaderless translation, novel translation initiation sites, and ribosome pausing at single codon resolution".*
- Uritskiy, G., Tisza, M.T., **Gelsinger, D.R.**, Munn, A., Taylor, J., and DiRuggiero, J. **2020**. "*Cellular life from the three domains and viruses are transcriptionally active in a hypersaline desert community*".

In preparation

- **Gelsinger, D.R.**, Reddy, R., Whittington, K., and DiRuggiero, J. **2020**. "*Post-transcriptional regulation of redox homeostasis during oxidative stress in haloarchaea by the small RNA SHOxi".*

AWARDS AND FELLOWSHIPS

2019	Travel award for invited speakers to the International Workshop on Geo-Omics of Archaea (IWGOA)
2019	Invited Discussion Leader for the Gordon Research Seminar Archaea
2017	Best poster award, Gordon Research Conference Archaea
2017	Trends in Microbiology best presentation award, Gordon Research Conference Archaea
2015	1st place award for JHU CMDB Annual Retreat Poster Competition, The Johns Hopkins University
2015	Institution Nominee for HHMI Gilliam Fellowship, The Johns Hopkins University
2014–2015	Wyche Fellowship, The Johns Hopkins University
2013	Instructional Related Activities (IRA) Grant, San Francisco State University
2013–2014	CSU Louis Stokes Alliance for Minority Participation (LSAMP), San Francisco State University
2013–2014	NIH Minority Access to Research Careers (MARC), San Francisco State University
2013 & 2014	NSF University of Nevada, Las Vegas-China PIRE Undergraduate Research Experience
2012	Institutional Nominee for HHMI Exceptional Research Opportunities Program (EXROP), San Francisco State University
2012–2013	NIH MBRS Research Initiative for Scientific Enhancement (RISE), San Francisco State University

PRESENTATIONS

September 2019 *International Workshop on Geo-Omics of Archaea (IWGOA), SUSTech University, Shenzhen, China*
Ribosome profiling reveals translation dynamics and alternative reading frame usage in Archaea (Invited speaker)

September 2019 *JHU CMDB Annual Research Retreat, Liberty Mountain, PN*
Ribosome profiling in Archaea reveals leaderless translation, novel translation initiation sites, and ribosome pausing at single codon resolution (poster)

July 2019 *Gordon Research Conferences Archaea: Ecology, Metabolism & Molecular Biology, Les Diablerets, Switzerland*
SHOxi: a small RNA post-transcriptional regulator in the oxidative stress response of *Haloferax volcanii* (Poster)

January 2019 *Mid-Atlantic Microbiome Meetup (M^3), Baltimore, Maryland*
Novel insight in identifying and monitoring regulatory small non-coding RNAs in halophilic microbial communities from the Atacama Desert (Poster)

October 2018 *JHU CMDB Annual Research Retreat, Rocky Gap, MD*
Profiling global translational effects and translation regulation in an oxidative stress-resistant haloarchaea (Talk)

March 2018 *5th Meeting of Regulating with RNA in Bacteria & Archaea, Seville, Barcelona*
Intergenic sRNA277.2: a post-transcriptional regulator of the oxidative stress response of the haloarchaeon *Haloferax volcanii* (Poster)

August 2017 *Gordon Research Conferences Archaea: Ecology, Metabolism & Molecular Biology, Waterville valley, New Hampshire*
Transcriptional landscape and regulatory roles of small non-coding RNAs in the oxidative stress response of the haloarchaeon *Haloferax volcanii* (Poster)

October 2015 *JHU CMDB Annual Research Retreat, Saint Michaels, Maryland*
Surviving in Extreme Environments by Regulating the Oxidative Stress Response in Haloarchaea (Poster)

July 2015 *Gordon Research Conferences Archaea: Ecology, Metabolism & Molecular Biology, Sunday River, Maine*
Small RNA-mediated Regulation of Stress Response in Haloarchaea (Invited Talk)

April 2014 *ASBMB Annual Meeting, San Diego, California*
Chemical inhibition of ammonia oxidation in the thermophilic nitrifying archaeon *Nitrosocaldus yellowstonii* (Poster)

January 2014 *CSUPERB Annual Meeting, Santa Clara, California*
Chemical inhibition of ammonia oxidation in the thermophilic nitrifying archaeon *Nitrosocaldus yellowstonii* (Poster)

June 2013 *Symposium: China-US Collaborative Research on Life in Terrestrial Geothermal Springs, Yunnan University, Kunming, China*
Inhibition of Ammonia Oxidation in the Thermophilic Ammonia-Oxidizing Archaeon *Nitrosocaldus yellowstonii* (Poster)

May 2013 *15th Annual COSE Student Project Showcase, San Francisco State University*
Inhibition of Ammonia Oxidation in the Thermophilic Ammonia-Oxidizing Archaeon *Nitrosocaldus yellowstonii* (Poster)

August 2012 *SFSU Undergraduate Summer Research Symposium, San Francisco State University*
Nitrosocaldus yellowstonii: Ammonia Oxidation at Low NH_3 Concentrations (Talk)

RELEVANT EXPERIENCE

Graduate Research Scholar – Johns Hopkins U. Microbial RNA biology, genomics, & ecology *2014–2020*

I use high throughput sequencing techniques (metagenomics/transcriptomics/ribosome profiling) to probe microbial dynamics of hypersaline endolithic communities in the Atacama Desert, Chile. I am using bioinformatic approaches to identify novel small non-coding RNAs that regulate stress response in haloarchaea found in these endolithic communities to confer adaptation to such extreme environments. I am

doing this using a two-pronged approach: (1) studying the regulation of the oxidative stress response of a cultured representative of the endolithic communities, *Haloferax volcanii*, in order to use genetics to elucidate a mechanistic actions of sRNAs, and (2) development of a bioinformatic program, SnapT (https://github.com/dgelsin/SnapT), to identify and track sRNAs *in situ* over temporal scales in the endolithic communities. After *in silico* analysis, results have been validated with *in vivo* and *in situ* experiments such as qPCR, northern blots, genetic manipulations (knockout, overexpression), and mRNA-sequencing. The majority of this work has focused on the most up-regulated sRNA during oxidative stress in *H. volcanii*. I found that this sRNA, SHOxi, is highly specific to oxidative stress conditions, that it is required for increased survival, and that its mechanistic function is to post-transcriptional degrade a messenger RNA involved in redox homeostasis through RNA-RNA interactions, effectively reducing the amount of damage during oxidative stress. This work has culminated into two publications (see PUBLICATIONS), and one more in preparation. Future work includes establishing the direct mechanistic action of SHOxi in *Haloferax volcanii* on all its potential targets and whether RNA-binding proteins are involved.

An alternative project I work on is the first highly resolved view of translation in Archaea. This involved the adaptation and development of ribosome profiling for an extremophile with 2-3M intracellular salt concentrations. Coupling ribosome profiling with inhibitors and stress conditions I found evidence that many genes not only initiate on alternative translation start sites (aTSS) around and within open reading frames (ORFs) but that these aTSS may be regulated translationally during stress. This work demonstrates how a microbe with a gene dense genome can potentially produce proteins with distinct functions (isoforms) using the same gene. This work is currently under review.

Undergraduate Research Scholar – San Francisco State U. Microbial ecology
2012–2014
I employed molecular, microbiological, and bioinformatic techniques to investigate the physiology and microbial interactions of nitrifying thermophilic archaea in geothermal springs. Major projects included: 1) transcriptional response to low and high substrate (NH_3) concentrations in AOA, 2) chemical inhibition of thermophilic AOA ammonia oxidation activity, and 3) exploring the physiology, genomics, and evolutionary relationship between novel thermophilic AOA from Chinese hot springs.

Summer Researcher Scholar – Tongji U. Shanghai, China, Biogeochemistry *August 2013 & 2014*
Employed organic chemistry and isotopic techniques to investigate carbon metabolism of AOA. Used a variety of ^{13}C-labeled carbon compounds in hot spring microcosms and extracted lipid biomarkers to use in conjunction with Liquid Chromatography-Mass Spectrometry and Gas Chromatography-Infrared-Mass Spectrometry.

FIELD WORK

Atacama Desert, Chile (2 trips) *January 2016 & February 2017*
In collaboration with the NASA ARADS team we tested the capabilities of new rover technology to drill and collect deep subsurface samples for microbial phylogenetic and functional analysis. In addition, I did independent research on the dynamic metabolic capabilities of the endolithic microbial communities occupying halite nodules using *in situ* RNA acquisition and sequencing approaches.

Tengchong, Yunnan, China
July 2013
Participated in a collaborative NSF research project, Partnerships in International Research and Education (PIRE), to assess the ammonia oxidizing archaeal (AOA) diversity and investigate their metabolism and physiology in high temperature hydrothermal environments.

PROFESSIONAL DEVELOPMENT WORKSHOPS

CMDB Bioinformatics Bootcamp, Johns Hopkins University
Fall 2014
Week intensive course on big data analysis, Python scripting, and biostatistics taught by Galaxy founder Dr. James Taylor.

Bioinformatics Programming Group, San Francisco State University
Fall 2013
Learned Unix and became familiar with regular expressions, scripting, Python, and R.

SYNERGISTIC ACTIVITIES (EDUCATIONAL AND OUTREACH)

2017	Designed and taught a lecture on microbial genomics and bioinformatics to undergraduates in the Microbiology course at the Johns Hopkins University Homewood Biology Department
2017	Laboratory Sustainability Representative of a pilot sustainability project at JHU to make laboratories more green and sustainable
2016–present	Lead member of the graduate student group Teachers and Researchers United (TRU) in order to promote better conditions for graduate students at JHU
2016	Invited and taught a bioinformatics lecture at the Halophiles 2016 meeting
2016	Invited and taught Quantitative Biology Bootcamp to Master Students at the University of Puerto Rico Mayaguez Campus
2016	Mentor for an undergraduate (Aldo Salazar Morales) recipient of the NSF REU
2016	Teacher's Assistant in Cell Biology
2015–present	Executive lead position of the Mentoring to Inspire Diversity in Science (MInDS) group at the Department of Biology, the Johns Hopkins University. MInDS is a peer-mentorship and outreach group for graduate students the in the CMDB program at JHU
2015	Mentor in STEM Achievement in Baltimore Elementary Schools (SABES)
2015	Teacher's Assistant in Biochemistry
2012–2014	Mentor in Biology Undergraduate Mentoring Program (BUMP) which partners successful upper division biology majors with entering biology majors to provide them with guidance and advice. Influential in shaping the direction of the program and bringing other students into research

STUDENT MENTORSHIP

Rahul Reddy – to date	Undergraduate student	09/2014
Kevin Maciuba	CMDB PhD rotation student	Fall 2015
Gherman Uritskiy	CMDB PhD rotation student	Spring 2016
Aldo Morales	REU undergraduate (Puerto Rico)	Summer 2016

Ariel Parker	CMDB PhD rotation student	Fall 2016
Neta Shwartz	CMDB PhD rotation student	Winter 2016
Katherine Huffer	CMDB PhD rotation student	Fall 2017
Michelle Biederman	CMDB PhD rotation student	Winter 2017
Anthony Mclean	CMDB PhD rotation student	Spring 2018
Kathleen Whittington	Undergraduate student	03/2018 – to date
Diego Tanton	Undergraduate student	03/2018 – 08/2018
Emma Dallon	CMDB PhD rotation student	Fall 2018
Grace Tamoefolau	CMDB PhD rotation student	Fall 2018
Sarah (Katie) Farney	CMDB PhD rotation student	Winter 2018
Sara Debic	CMDB PhD rotation student	Fall 2019
Dylan Taylor	CMDB PhD rotation student	Winter 2020
Rima Sakhawala	CMDB PhD rotation student	Winter 2020

LANGUAGES

Fluency in Spanish (mother tongue), English, and French (advanced to intermediate).